Supplementary Materials for

Structural basis for membrane insertion by the human ER membrane protein complex

Tino Pleiner*, Giovani Pinton Tomaleri*, Kurt Januszyk*, Alison J. Inglis, Masami Hazu, Rebecca M. Voorhees†

*These authors contributed equally to this work.
†Corresponding author. Email: voorhees@caltech.edu

Published 21 May 2020 on *Science* First Release
DOI: 10.1126/science.abb5008

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at science.sciencemag.org/cgi/content/full/science.abb5008/DC1)

MDAR Reproducibility Checklist (.pdf)
Movies S1 to S3 (.mp4)

Materials and Methods

Plasmids and antibodies

The plasmid for generating the nanodisc-forming protein MSP2N2 was acquired from Addgene (ID 29520). Expression constructs for the *bd*SENP1 and *bd*NEDP1 proteases (Addgene IDs 104962 and 104129) (*31*, *32*), as well as the $SUMO^{Eu1}$ module and its cognate $SENP^{EuB}$ protease (Addgene ID 149333) (*33*) were kind gifts from Dirk Görlich. 2nd generation lenti-viral packaging and envelope plasmids were kind gifts from Carlos Lois.

The following antibodies were used in this study: EMC2 (25443-1-AP, Proteintech, USA); EMC4 (27708-1-AP, Proteintech, USA); EMC6 (ab84902, Abcam, UK); EMC8 (STJ117038, St. John's Laboratory, UK); Anti-FLAG-HRP (Millipore-Sigma, USA); Anti-HA (Millipore-Sigma, USA); Anti-3F4, Anti-GFP, and Anti-BAG6 were kind gifts from Ramanujan Hegde and have been described previously (*34*).

E. coli protein expression and purification

Untagged GFP used as a negative control in fig. S9 was expressed with an N-terminal His_{14}-*bd*NEDD8 tag (*31*) in *E. coli* NEB express I^q (New England Biolabs, USA). Expression was carried out at a 1 L scale in SuperBroth medium for 6 hours at 18˚C using 0.2 mM IPTG for induction. Harvested cells were resuspended in lysis buffer (50 mM Tris/HCl pH 7.5, 300 mM NaCl, 20 mM imidazole, 1 mM DTT, 1 mM PMSF) and lysed by sonication (4x 1 min pulsing in thin-walled metal containers in an ice-water bath, Macro-Tip at 100% amplitude, Branson Sonifier). GFP was purified from lysate by Ni^{2+}-chelate affinity chromatography and eluted with protease elution buffer (300 nM *bd*NEDP1 in 50 mM Tris/HCl pH 7.5, 300 mM NaCl, 20 mM imidazole, 250 mM sucrose) to remove the N-terminal tag.

The expression and purification of GST-3xFLAG-Calmodulin and BpA-RS used for in vitro crosslinking (fig. S9) was carried out as described previously (*35*). Expression and purification of the MSP2N2 protein was achieved as described by established protocols (*36*).

Expression and purification of biotinylated anti-GFP nanobody

An *E. coli* expression vector (pTP396, Addgene ID 149336) encoding the anti-GFP nanobody "GBP1/Enhancer" (*37*) fused to an N-terminal His_{14}-Biotin acceptor peptide (Avi)-$SUMO^{Eu1}$-tag was generated (*14*, *33*). The Avi-tag is a 15 amino acid peptide with a single lysine that can be specifically biotinylated using purified *E. coli* biotin ligase BirA (*38*, *39*). The engineered $SUMO^{Eu1}$ module is resistant to cleavage by endogenous deSUMOylases in eukaryotic cell lysates, enabling stable isolation of nanobody:target complexes from eukaryotic extracts (*14*, *33*). The corresponding engineered $SENP^{EuB}$ protease cleaves $SUMO^{Eu1}$ very efficiently at low concentration on ice. Alternatively, a commercial system can be used, based on the SUMOStar tag and SUMOStar protease (LifeSensors, USA) (*40*).

The nanobody fusion protein was expressed in *E. coli* NEB Express I^q cells. A 100 ml SuperBroth pre-culture was grown overnight at 28˚C and diluted to 1 L after ~18 hours. After dilution, the culture was incubated at 18˚C until an OD_{600} of ~2.0 (2-4 hours). Expression was then induced by addition of 0.2 mM IPTG for 18-20 hours at 18˚C. Cells were harvested and the pellet was resuspended in 120 ml lysis buffer (50 mM Tris/HCl pH 7.5, 300 mM NaCl, 20 mM imidazole, 1 mM DTT, 1 mM PMSF) and lysed by freeze-thaw followed by sonication (4x 1 min pulsing in thin-walled metal containers in an ice-water bath, Macro-Tip at 100% amplitude, Branson Sonifier). The lysate was cleared by centrifugation for 45 min at 17,000 rpm and 4˚C in an SS-34 rotor. Half of the lysate was frozen after addition of 250 mM sucrose for later purifications. The other half (60 ml) was incubated in batch with ~4 ml bed volume Ni^{2+}-NTA agarose (Thermo Fisher Scientific, USA) for 1 hour at 4˚C and then transferred to a gravity flow column. The resin was washed with three column volumes of lysis buffer before stepwise elution with imidazole elution buffer (50 mM Tris/HCl pH 7.5, 300 mM NaCl, 500 mM imidazole, 1 mM DTT, 250 mM sucrose). The buffer of the eluate was then exchanged to storage buffer (50 mM Tris/HCl pH 7.5, 200 mM NaCl, 1 mM DTT and 250 mM sucrose) using a PD-10 desalting column (GE Healthcare, USA). The purified protein was aliquoted, snap-frozen in liquid nitrogen, and stored at -80˚C until further use.

Purification of *E. coli* biotin ligase BirA and nanobody biotinylation

The expression of biotin ligase BirA from *E. coli* and its use for biotinylation of Avi-tagged proteins has been described extensively before (*38*, *39*) (commercial products by Avidity, USA). His_{14}-*bd*NEDD8-tagged BirA Biotin ligase was expressed from pTP264 (Addgene ID 149334) in *E. coli* NEB express I^q (New England Biolabs, USA) for 18-20 hours at 18˚C in 1 L SuperBroth using 0.2 mM IPTG for induction. After cell harvest, lysis and binding to Ni^{2+}-resin as described above, tagged BirA was eluted with imidazole elution buffer. The buffer was exchanged to storage buffer as described above. The purified protein was aliquoted, snap-frozen in liquid nitrogen, and stored at -80˚C until further use.

An in-solution biotinylation reaction contained 300 µl 5x biotinylation buffer (250 mM Tris/HCl, 500 mM NaCl, 62.5 mM $MgCl_2$, 50 mM ATP, 50 mM biotin), at least 50 µM purified Avi-tagged anti-GFP nanobody (from pTP396), a 1:50 molar ratio of purified His-tagged BirA, and water to add up to 1.5 ml final volume. The reaction was incubated for 3 hours at 25˚C and then applied to a PD-10 desalting column (GE Healthcare, USA) equilibrated in storage buffer to remove excess biotin. Fractions with normal 260/280 ratio (comparable to starting ratio of the prep, usually ~0.6-0.7) were pooled and quantitative biotinylation was assessed by test binding to Streptavidin beads. If fully biotinylated and added below bead capacity, all of the nanobody should be in the bound fraction and almost none should remain in the unbound fraction.

Expression of *bd*$SENP^{EuB}$ protease

His_{14}-Tev-tagged *bd*$SENP^{EuB}$ protease (Addgene ID 149333) (*33*) was expressed in *E. coli* NEB express I^q as described in the '*E. coli* protein expression and purification' section and purified by Ni^{2+}-chelate affinity chromatography using imidazole elution. The buffer of the eluate was then exchanged to storage buffer as described above.

Mammalian in vitro translation

Cell free in vitro translations were carried out in rabbit reticulocyte lysate (RRL) as previously described (*41*). Constructs for expression in RRL were based on the SP64 vector (Promega, USA). Briefly, templates for in vitro transcription were prepared by PCR and contained a 5' SP6 promoter, an in-frame stop codon, and a short 3' UTR. PCR products were purified and used for in vitro transcription (1.5 hours at 37˚C) and translation (20-30 min at 32˚C) in the presence of radioactive ^{35}S-methionine as described (*41*). Translation of membrane subunits was performed in RRL supplemented with nucleased canine pancreatic microsomes (cRM), prepared as previously described (*42*), at a ratio of 0.5 µl of cRM per 10 µl translation reaction. Native immunoprecipitations were carried out for soluble complexes in physiologic salt buffer (PSB) (50 mM HEPES pH 7.5, 130 mM KAc, 2 mM $MgAc_2$, 1 mM DTT) using Anti-FLAG M2 affinity resin or Anti-HA Agarose (both Millipore-Sigma, USA).

Immunoprecipitation of membrane subunits was carried out as follows. Following the translation reaction, membranes were isolated by centrifugation (20 min at 55,000 rpm in a Beckman TLA55 rotor) through a 25% (w/v) sucrose cushion in PSB. Following resuspension, complexes were solubilized in 1.75% (w/v) digitonin in PSB on ice for 10 min. Insoluble material was removed by centrifugation for 15 min at 20,000 x g, and the supernatant was incubated with the appropriate resin for 1.5 hours at 4˚C. The resin was washed with PSB supplemented with 0.25% (w/v) digitonin, and eluted with either 3xFLAG or 1xHA peptide. Complexes were analyzed by SDS-PAGE and autoradiography.

Protease digestions were performed for one hour on ice by the addition of 0.5 mg/ml proteinase K to translation reactions. The digestion was quenched by the addition of 5 mM PMSF in DMSO, followed by transfer to boiling 1% (w/v) SDS in 0.1 M Tris/HCl pH 8.0 (pH determined at room temperature). Immunoprecipitation of protected fragments was performed in IP buffer (50 mM HEPES/KOH pH 7.5, 100 mM KAc, 2 mM $MgAc_2$, and 1% (v/v) Triton X-100) using the 3F4 or HA antibody and protein G resin. Samples were subsequently analyzed by SDS-PAGE and autoradiography.

In vitro translation in the PURE System and photocrosslinking

Site-specific UV crosslinking was carried out as previously described (*35*). Constructs for translation in the PURE system were derived from the T7 promotor-based PURExpress plasmid provided by New England Biolabs (USA). Human Sec61β with residue Phe85 in its transmembrane helix mutated to an amber stop codon, was translated in the presence of ^{35}S-methionine, 100 nM $CaCl_2$, and 10 µM purified Calmodulin (CaM) in the coupled transcription/translation PURE system (New England Biolabs, USA). This translation extract contains a defined set of purified *E. coli* ribosomes and translation factors and lacks all release factors (ΔRF). The release factors RF2 and RF3, but not RF1 (which recognizes the UGA [amber] stop codon) were added back to the reaction. The unnatural amino acid BpA (100 µM) was incorporated at the amber stop codon using recombinant BpA synthetase (100 µg/ml) and suppressor tRNA purified as described (*35*). After translation for 2 hours at 32˚C and addition of 1 mM puromycin, the reaction was layered

on top of a 20% (w/v) sucrose cushion prepared in PSB with 100 nM $CaCl_2$ and spun for 1 hour at 55,000 rpm in a TLS-55 rotor (Beckman Coulter, USA) to remove aggregates. Sec61β-CaM complexes were retrieved from the cushion and incubated in the presence of 1 mM EGTA with either buffer, purified GFP, or EMC purified via EMC5-GFP or EMC3-GFP from the respective stable HEK293 cell lines (described below). Except for the -UV control lane, all reactions were irradiated at a distance of ~7-10 cm with a UVP B-100 series lamp (Analytik Jena, Germany) for 15 min on ice, before quenching with SDS-PAGE sample buffer. Samples were analyzed by SDS-PAGE and autoradiography. Comparable crosslinking efficiency was observed in at least ten independent replicates. A representative gel is shown in fig. S9.

Generation of stable HEK293 cell lines

Flp-In 293 T-Rex cells were purchased from Thermo Fisher Scientific (USA) (RRID: CVCL_U427). Cell lines were grown in DMEM supplemented with 2 mM glutamine, 10% (w/v) FBS, 15 µg/ml Blasticidine S, and 100 µg/ml Zeocin. The open-reading frame to be integrated into the genomic FRT site was cloned into the pcDNA5/FRT/TO vector backbone and cell lines were generated according to the manufacturer's protocol. Briefly, the open-reading frame was cloned into the pcDNA5 backbone and transfected together with pOG44 Flp-In recombinase in a 9:1 ratio using Trans-IT 293 transfection reagent (Mirus, USA) according to the manufacturer's instructions. 48 hours after transfection, 100 µg/ml Hygromycin B was added to select for cells that had undergone successful integration. This procedure was used to generate cell lines stably expressing GFP-EMC2-2A-RFP, EMC5-GFP-2A-RFP, EMC3-GFP-2A-RFP, OPRK1-GFP-2A-RFP, TRAM2-GFP-2A-RFP. Constructs for generating cell lines expressing the RFP-tagged transmembrane helix and flanking regions of human squalene synthase (SQS/FDFT1) or vesicle-associated membrane protein 2 (VAMP2) were prepared as previously described (*3*).

To facilitate large-scale growths, the adherent cell line expressing GFP-EMC2 was adapted to grow in suspension. This was achieved over the course of ~10 days, during which the FBS-supplemented DMEM was serially diluted with FreeStyle 293 Expression Medium (Thermo Fisher Scientific, USA). Once growing in 100% FreeStyle Medium, the cells were transferred to 1-2 L roller bottles (Celltreat, USA) and grown in a shaking incubator operating at 8% CO_2 and rotating at 125 rpm.

Flow cytometry analysis of reporter cell lines

To measure protein stability in cells, stable cell lines were generated as described above containing a fusion of a protein-of-interest with either GFP or RFP in a GFP-2A-RFP backbone (*43*). Note, the mCherry sequence is used throughout, but is referred to as RFP for simplicity. The viral 2A sequence induces skipping of a peptide bond by the ribosome (*44*), resulting in two separate polypeptide chains being produced from a single transcript in a 1:1 ratio. The GFP and RFP fluorescence intensity can be measured by flow cytometry to derive a GFP:RFP or RFP:GFP ratio. Any changes in the GFP:RFP ratio (for OPRK1 and TRAM2) or RFP:GFP ratio (for SQS and VAMP2) reflect changes in the post-translational stability of the fused substrate or control protein.

To test the effect of EMC subunit mutations on EMC function in cells, lenti-viral vectors in a pHAGE2 backbone were generated that stably integrate and express either wild type or mutant subunit from a CMV promotor-driven open-reading frame. EMC3 variants were fused to a C-terminal TagBFP-3xFLAG tag, while EMC6 variants carried a C-terminal 3xHA tag followed by a 2A sequence and TagBFP. A 2nd generation packaging system was used to generate lenti-viral particles. Reporter cell lines expressing GFP or RFP-fused EMC substrates or control proteins were then transduced with wild type or mutant EMC subunit lenti-viral vectors. BFP-positive cells were selected for analysis of the effect of the mutants on the GFP:RFP or RFP:GFP ratio of the underlying reporter. The effect of mutation of the following residues was tested in at least three independent replicates in all cell lines: EMC3 $M^{Cyt\text{-}1}$-S, R31 and R180. Representative data is shown.

In order to verify incorporation of mutant EMC subunits into the intact EMC, an aliquot of the analyzed cells was resuspended in solubilization buffer (50 mM HEPES/KOH pH 7.5, 200 mM NaCl, 2 mM $MgAc_2$, 1% (w/v) DDM [Anatrace, USA], 1 mM DTT, 1x complete EDTA-free protease inhibitor cocktail [Roche, Germany]). After incubation on ice for 30 min, the lysate was centrifuged for 10 min at 18,000 x g at 4˚C in a table-top centrifuge. The cleared detergent lysate was then subjected to FLAG-IP and elution with 3xFLAG peptide for EMC3 or HA-IP and elution with 1xHA peptide for EMC6. Eluates were analyzed by SDS-PAGE and Western blotting.

Purification of EMC for structure determination

GFP-tagged proteins were purified using a previously described native single-step strategy (*14*) (see fig. S1B, C). It relies on a multifunctional tag that allows biotinylation and immobilization of nanobodies onto magnetic Streptavidin beads. A specific protease cleavage module in the fusion tag further enables gentle release of the nanobody along with its bound target complex on ice in physiological buffer, while avoiding elution of unspecific background binders. See figs. S1C and S9B for representative examples. Magnetic beads are preferred over porous beads, as they show higher capacity for larger protein complexes and also enable elution in a very small volume.

2 L of HEK293 cells stably expressing GFP-EMC2-2A-RFP were grown in suspension and induced for 48 hours with 1 µg/ml doxycycline (DOX). Cells were harvested, washed with 1x PBS, weighed and resuspended with 6.8 ml solubilization buffer per 1 g cell pellet (50 mM HEPES/KOH pH 7.5, 200 mM NaCl, 2 mM $MgAc_2$, 1% (w/v) DDM [Anatrace, USA], 1 mM DTT, 1x complete EDTA-free protease inhibitor cocktail [Roche, Germany]). After 1 hour of head-over-tail incubation with solubilization buffer, the lysate was cleared by centrifugation for 40 min at 4˚C and 30,000 x g in a Sorvall RC6+ centrifuge. In parallel, beads with immobilized nanobody were prepared similarly as described before (*14*). Briefly, 60 µl resuspended Pierce Streptavidin magnetic beads (Thermo Fisher Scientific, USA) per 1 g cell pellet were pre-equilibrated in wash buffer (solubilization buffer with 0.015% (w/v) DDM). Then 20 µg biotinylated His_{14}-Avi-$SUMO^{Eu1}$-tagged anti-GFP nanobody (see above for expression protocol) per 60 µl beads were immobilized for 20 min on ice with occasional mixing. After this, all remaining biotin binding sites were blocked by incubation with 100 µM biotin in 50 mM HEPES/KOH pH 7.5 for 5 min on ice with occasional mixing. The beads are then washed

with wash buffer and incubated with the cleared detergent cell lysate for 1 hour, binding head-over-tail at 4˚C. The beads were washed four times with 1 ml wash buffer and finally, the anti-GFP nanobody along with all bound proteins was released under native conditions and in minimal volume (~20-30 µl for 60 µl beads) by cleavage with 250 nM $SENP^{EuB}$ protease (*33*) in wash buffer for 30 min at 4˚C with occasional mixing.

The eluted EMC was reconstituted into lipid nanodiscs as follows. His_6-tagged MSP2N2 (Addgene #29520) was expressed in *E. coli* BL21(DE3) cells and purified as previously described (*36*). DDM solubilized EMC at ~5 µM concentration was incubated on ice with POPC 16:0-18:1 lipids (Avanti, USA) for 30 min, after which MSP2N2 was added to a final ratio of EMC:MSP2N2:POPC of 1:4:300. The reaction was incubated for 2 hours on ice. Bio-Beads SM-2 (Bio-Rad Laboratories, USA) were washed sequentially in methanol, water, and finally equilibrated in wash buffer without detergent. The prepared Bio-Beads were added to the EMC:MSP2N2:POPC mixture and incubated for 18 hours at 4˚C with continuous shaking. Bio-Beads were removed by centrifugation, and the reconstituted complex was purified via size exclusion chromatography on a 3.5 ml Superose 6 column (GE Lifesciences, USA) (fig. S1). The purified reconstituted complex was concentrated to 0.2 mg/ml using an Amicon Ultra 0.5 ml 100K MWCO concentrator (Millipore-Sigma, USA), and 0.05% (w/v) CHAPSO (Millipore-Sigma, USA) was added immediately prior to vitrification.

Grid preparation and data collection

3 µl of sample was applied to Holey carbon grids (Quantifoil R1.2/1.3) that were glow discharged using a Pelco easiGlow, Emeritech K100X at a plasma current of 20 mA for 60 seconds in air. The grid was blotted at 6˚C, 95% humidity with filter paper (Whatman No.1) for 3-5 seconds at a -4 blot force before plunging into liquid ethane using the FEI Vitrobot Mark v4 x2. Data were collected on an FEI Titan Krios equipped with an energy filter (20 eV slit width) operated at 300 keV and a K3 (Gatan) direct detector. Images were recorded using an automated acquisition pipeline in SerialEM (*45*). Illumination conditions were adjusted in nanoprobe mode to a fluence of 13 e^-/pixel/second. Images were recorded between -2.0 and -0.5 µm defocus in super resolution mode at a calibrated magnification of 0.418 Å/pixel. Two-second images with a frame width of 50 ms (1.48 e^-/$Å^2$/frame) were collected in counting mode.

Image processing

6,345 movies were processed using cryoSPARC v2.13.2 to generate dose-weighted aligned and averaged micrographs alongside aligned movies (*46*). During motion correction, movies were downsampled two-fold to 0.836 Å/pixel and all subsequent processing was carried out at this pixel size. Of the 6,345 movies acquired, a sub-set of 6,277 movies were selected for further processing based on an estimated resolution better than 4.5 Å. Using the automated Blob Picker function, particles were picked and then manually filtered to remove obvious debris. The remaining particle stack of 440,033 particles was further filtered by an initial round of 2D classification to remove junk particles (fig. S2C). The resulting 66,137 particles were then used in cryoSPARC to generate six ab initio 3D reconstructions. Only one class was consistent in size with the EMC in a nanodisc (27,313 particles), while the other classes had very small sizes more

consistent with broken complexes or empty nanodisc. The putative EMC-nanodisc class was then refined, new templates were generated from the initial 3D model, and particles were re-picked to generate a new more homogeneous particle stack of 1,034,250 particles. Multiple rounds of 3D heterogeneous classification were employed to remove junk particles. A final round of classification was employed on a population of 213,610 particles that produced the 188,746 particles that were used for structural determination. Upon analysis of the EM density maps, it became clear that there was weak density that flanked one side of the putative TM region. Therefore, an additional round of 3D heterogeneous classification/refinement was performed that produced a population of 42,506 particles that had stronger EM density within this region with a worse apparent overall resolution (fig. S7).

The 188,746 pool of particles were exported from cryoSPARC using the csparc2star.py (UCSF pyem v0.5. Zenodo) suite of conversion scripts for additional processing in RELION-3.0 (fig. S2D) (*47*). Following an initial 3D refinement, CTF refinement was performed, resulting in an overall EM density map to 3.4 Å resolution (fig. S3). To address the apparent flexibility between the lumenal, membrane, and cytoplasmic domains, Multibody refinement in RELION-3.0 was performed using two different masking strategies. First, to better elucidate the lumenal domain, Multibody was performed using two bodies: (i) the lumenal domain and a portion of the transmembrane region, and (ii) the cytoplasmic domain and the remaining portion of the membrane region. Second, to better elucidate the cytoplasmic domain, Multibody was performed again with two bodies: (i) the membrane and cytoplasmic region, and (ii) the lumenal domain alone. Together these resulted in EM density maps for the lumenal region composed primarily of EMC1 to an overall resolution of 3.2 Å, and the cytoplasmic region composed of EMC2, 8 and the cytoplasmic portions of EMC3 and 5 to an overall resolution of 3.6 Å (fig. S3). Post-processing was performed with a soft-mask of 6 pixels, extended by 10 pixels, and the B-factor was estimated automatically in RELION-3.0 following standard procedures. Local resolution was calculated using RELION-3.0.

TM assignment

Two orthogonal strategies were used to assign the observed TMs. First, segmented maps were generated for each TM in Chimera (*48*). Based on the observed side chain density, an initial PDB was built using *phenix.map_to_model, phenix.map_to_sequence* (*49*), or manual building based on an idealized helix in COOT (*50*, *51*). This PDB was then input into CryoID (*52*) to search against the predicted sequences of all twelve EMC TMs, with an additional five flanking residues included on either side. The top-scoring hits were manually inspected in the map for assignment.

In the second strategy, TMs with clear connectivity in either the cytosol or the lumen were manually identified. The ‘overall map’ was also segmented in Chimera to generate individual maps for these TMs and their connecting loops. A putative assignment was made based on the predicted topology of each EMC subunit, and the sequence and segmented maps were used in *phenix.map_to_model* (*49*). The initial fitted TMs and loops were visually inspected and further refined by rounds of manual building in COOT (*50*, *51*) as described in Table S2, to generate a starting PDB.

These two orthogonal strategies converged on a single TM assignment, allowing for the confident assignment of eight of the well-ordered TMs (Fig. 1C-D and fig. S4A-C). However, the final ninth TM could not be assigned to any of the predicted subunit TMs even after extensive *phenix.map_to_model* (*49*) was carried out using the segmented unassigned TM map in combination to all unassigned TM sequence. Careful inspection of the EM density map revealed continuous density from residue 49 of EMC6 to an additional membrane-spanning region (fig. S6D). After careful manual building in COOT followed by *phenix.real_space_refine* (*53*) residues 12 to 48 of EMC6 could be modeled in this density. Proteinase K digestion analysis confirmed that indeed, EMC6 contains three TMs when translated in the presence of EMC5 (fig. S6A-C).

The density for the three TMs of EMC4, 7, and 10 was poorly ordered, and putatively assigned based on connections to their structured lumenal domains (fig. S7). Using the sub-population of 42,506 particles (described above), stronger EM density was detected within this region, and could be used to successfully model a backbone through the contiguous density that protruded from the termini of EMC4, 7, and 10. However, this region could not be unambiguously assigned and therefore these TMs were not included in the final deposited model.

Model building and refinement

Initial models for all of the EMC soluble domains were generated using the transform-restrained Rosetta (trRosetta) algorithm, a deep learning-based modeling method based on the Rosetta energy minimization pipeline with additional distance and interaction restraints generated from co-evolution (*54*).

Initially, the individual maps were inspected (cytosolic and lumenal), and each domain was individually fit each using the *phenix.dock_in_map* function of the PHENIX Software suite (*55*). The initial fitted models were visually inspected and further refined by iterative rounds of de novo model building in COOT (*50*, *51*) as described in Table S2.

Initial attempts to build regions that differed from the trRosetta model, including connecting loops, and the N- and C-termini of most subunits, were made by first segmenting the map in Chimera (*48*) for use in *phenix.map_to_model* (*49*). Subsequent manual building was performed in COOT (*50*, *51*), using secondary structure prediction generated by trRosetta and PSIPRED (*54*, *56*).

The resolution of maps for the lumenal region was in the range of 2.5 - 4.0 Å (fig. S3), allowing unambiguous model building. The presence of glycosylation sites at Asn residues (370, 818 and 913 of EMC1 and 182 of EMC10) and disulfide bond pairs (227-237 and 338-368 of EMC1) (fig. S8), validated the register of the polypeptide chain. For the cytosolic domain, the register was confirmed by the presence of large aromatic side chains and using structure-based mutagenesis (Fig. 2 and fig. S8).

The density for the flexible linkers between the top ß-propeller and lumenal helix of EMC1 (amino acids 475-478), the α-helix 14 and α-helix 15 of EMC2 (amino acids 276-279), the TM1 and the coiled-coil of EMC3 (amino acids 40-43), and between the TM1 and TM2 of EMC5 (amino acids 37-39) was not sufficient to unambiguously assign the position of the side chains, thus they are included in the model as poly-Ala. Similarly, the density for the lumenal region of EMC4 that forms a β-strand and completes one blade of the bottom β-propeller of EMC1, was tentatively assigned as P170 to L183 of EMC4 (fig. S8B). However, it was not sufficient to unambiguously assign the registry; therefore, in the model, the region is assigned as unknown amino acid (UNK) poly-Ala/Gly.

Initially, models for each region of the complex (TMs, cytosol and lumen) were refined against the respective EM density maps (overall, cytosolic and lumenal) using *phenix.real_space_refinement* (*53*) and ISOLDE (*57*). Secondary structure restraints were generated in PHENIX (*55*) and were maintained during the refinement. For the lumenal region, a set of restraints for N-acetyl-D-glucosamine (NAG) were generated using *phenix.elbow* (*58*). Finally, all models were assembled and refined against the 'overall map'. The FSC curves between individual models and full maps (overall, cytosolic and lumenal) as well as per-residue B factors and cross-correlation were calculated after the refinement and plotted in fig. S3 and S5. Statistics of the map reconstruction and model refinement are reported in Table S1. The final models were evaluated using MolProbity (*59*). All figures and movies were generated using Pymol (www.pymol.org) and Chimera (*48*). Subunit interface analyses were performed using the PDBsum (*30*) and PDBePISA server (*29*).

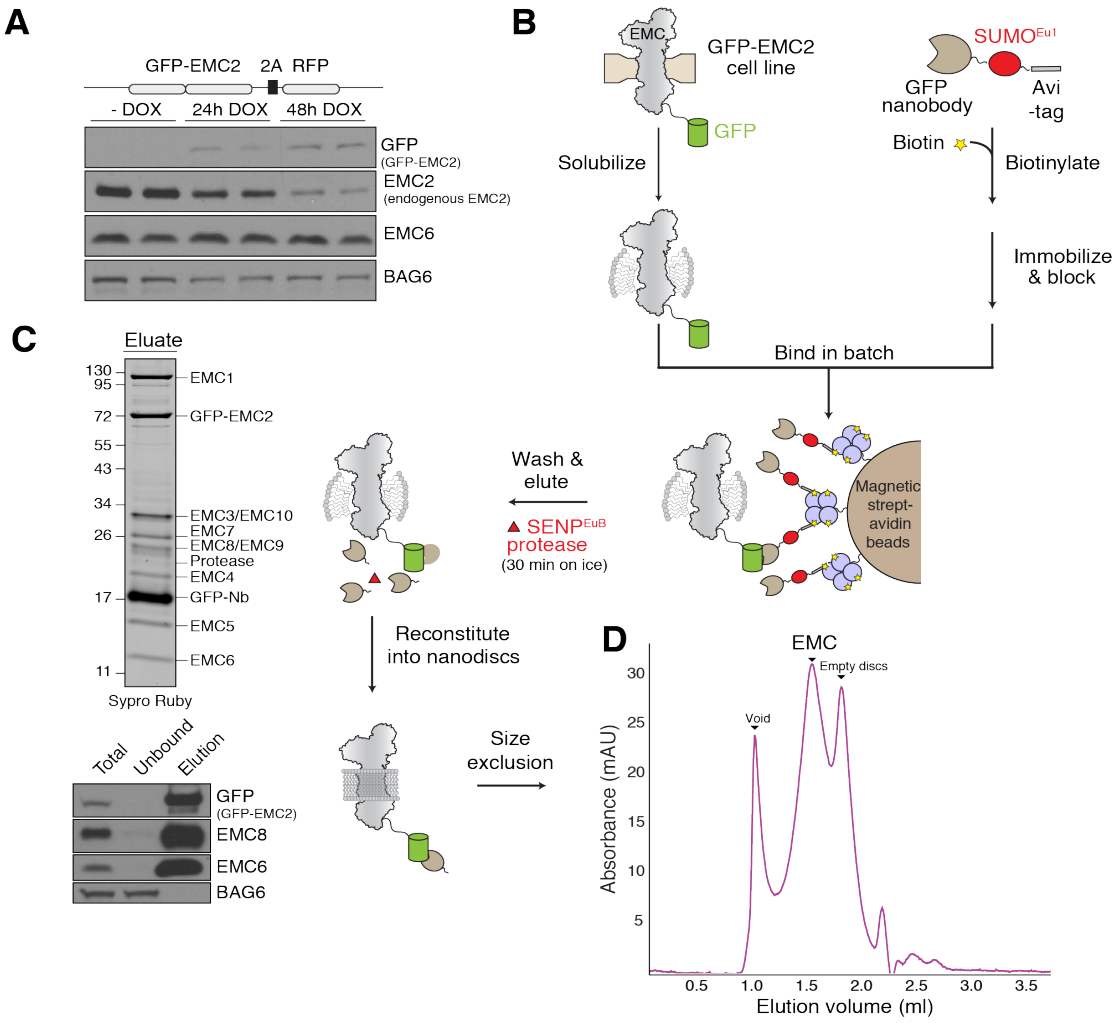


Fig. S1. Purification and nanodisc reconstitution of the human EMC.
(**A**) HEK293 cell lines stably expressing GFP-EMC2-P2A-RFP, where '2A' indicates the viral P2A sequence that causes peptide bond skipping by the ribosome, were harvested either 24 or 48 hours after treatment with carrier or doxycyline (DOX) and analyzed by Western blotting with the indicated antibodies. GFP-tagged EMC2 replaces its endogenous counterpart, which is largely degraded. (**B**) Schematic of the GFP-nanobody affinity purification strategy. HEK293 cells stably expressing GFP-EMC2 were adapted to grow in suspension to facilitate large-scale growths. GFP-tagged EMC2 was efficiently incorporated into the EMC, resulting in intact EMC fused to GFP and expressed at endogenous levels. In parallel an Avi tag-SUMO^{Eu1}-GFP-nanobody fusion was purified from *E. coli*, biotinylated using BirA, and immobilized on magnetic Streptavidin beads. Detergent solubilized EMC was affinity purified via the GFP tag, and specifically eluted from the resin under native conditions using SENP^{EuB} protease cleavage. All constructs needed to implement this strategy are available via Addgene (see Methods). (**C**) Top: SDS-PAGE gel of the purified EMC, stained with Sypro Ruby. GFP-tagged EMC2 is efficiently incorporated into the intact EMC and thus co-purifies all endogenous EMC subunits, which were identified by mass spectrometry. The native single-step purification described in (B) yielded EMC at sufficient quantities and purity to be utilized directly for

reconstitution into nanodiscs. Bottom: A fraction of the anti-GFP nanobody IP eluates was analyzed by Western blotting with the indicated antibodies. (**D**) Size exclusion chromatography of the DDM solubilized EMC reconstituted into lipid nanodiscs using the MSP2N2 scaffold and POPC lipid. UV absorbance at 280 nm was monitored during the purification to identify fractions containing the reconstituted complex.

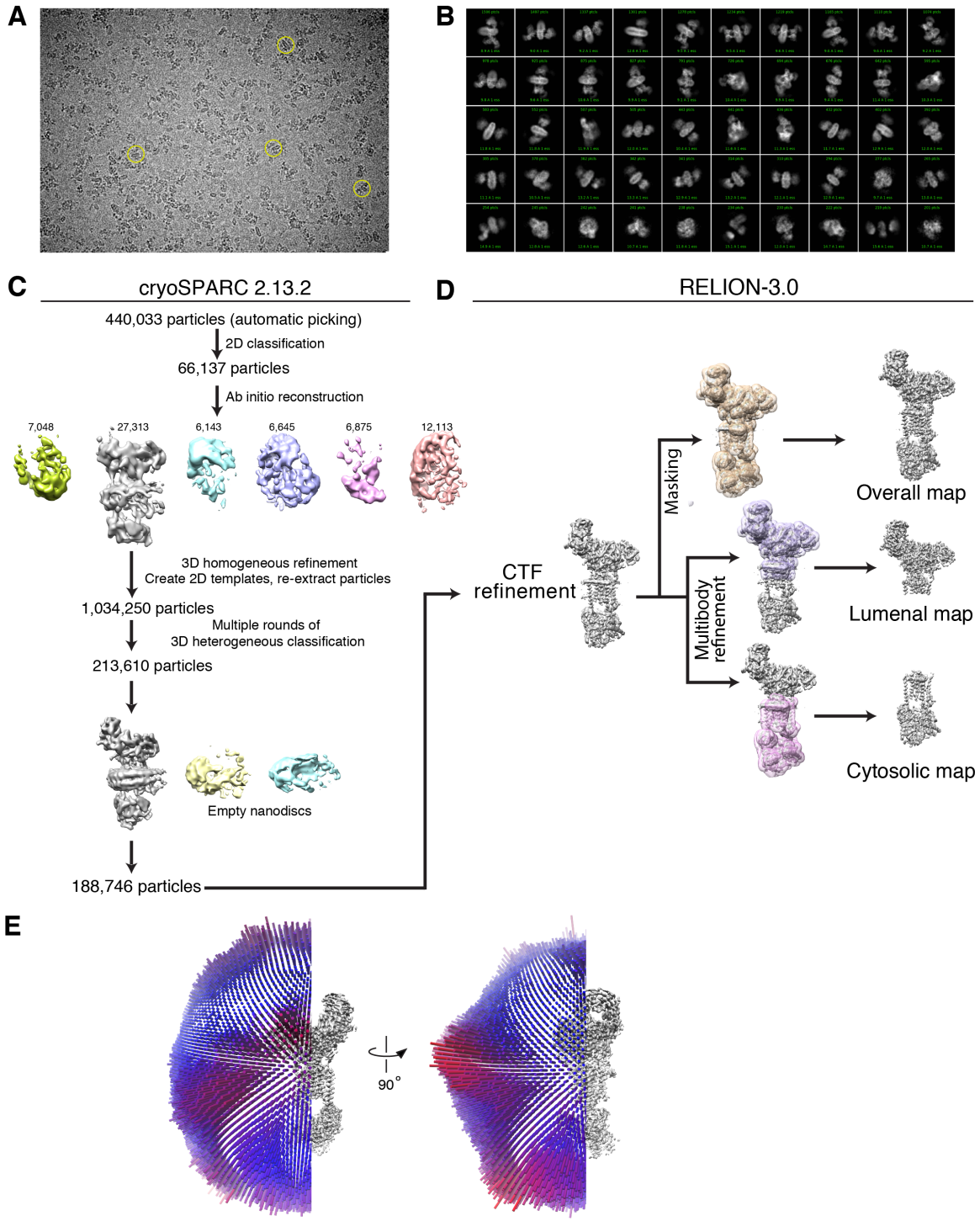


Fig. S2. Classification and refinement procedure.
(**A**) A representative image with several particles circled in yellow. A total of 6,345 images were collected, and 6,277 were used for further processing based on an estimated resolution better than 4.5 Å. (**B**) Representative 2D class averages after selecting and filtering for intact particles. (**C**) Classification scheme in cryoSPARC 2.13.2 used to identify a population of particles containing the intact EMC in a lipid nanodisc.

Following an initial round of automatic particle picking, iterative rounds of 2D classification and template-based particle picking were used to filter out the majority of empty nanodiscs and broken complexes. After re-extraction of 1,034,250 particles, multiple rounds of 3D heterogeneous classification were employed to characterize a sub-population of 213,610 particles; 188,746 of these particles appeared to contain intact EMC in a lipid nanodisc. (**D**) These 188,746 particles were exported from cryoSPARC for further refinement in RELION-3.0. Following CTF refinement, a mask that excluded the nanodisc was applied during 3D refinement resulting in an improvement in the overall resolution of the complex and a final 'overall map' used primarily for assignment and building of the TMs. Multibody refinement was used to generate improved maps for the cytosolic and lumenal domains, which allowed de novo modeling of these regions. This would not have been possible using the 'overall map'. (**E**) Orthogonal views of the angular distribution of particle views for the final data set (188,746 particles). Each cylinder represents one view. The number of particles with their particular orientations are represented by both length and color (from blue to red) of the cylinders.

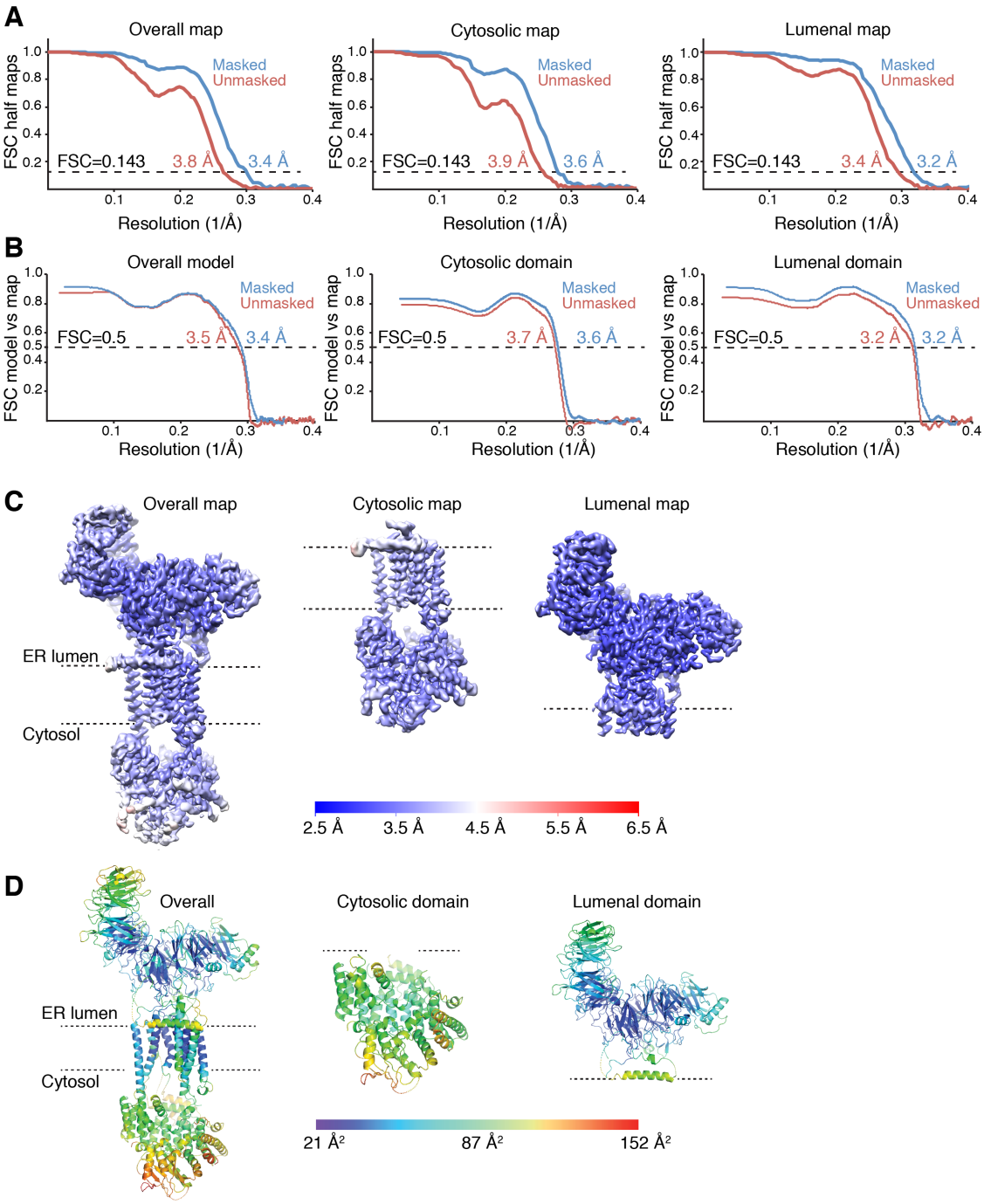


Fig. S3. Map and model quality.

(**A**) Gold-standard Fourier Shell Correlation (FSC) curves for the masked and unmasked half maps used for building of the EMC calculated using RELION-3.0. While the 'overall map' was sufficient for assigning and building the TMs, Multibody refinement in RELION-3.0 was able to correct for intersubunit flexibility to significantly improve the local resolution and molecular detail visible in the 'cytosolic' and 'lumenal' maps. The nominal resolution of each map is demarcated using the FSC = 0.143 criterion. (**B**) FSC curves of the model vs. map, masked and unmasked calculated using *phenix.real_space_refine*. Overall is the complete model vs. the 'overall map', cytosol is

the model of the cytosolic domain shown in (D) vs. the 'cytosolic map', and lumen is the lumenal domain vs. the 'lumenal map'. FSC = 0.5 is indicated for the masked and unmasked maps. (**C**) View of the final unsharpened cryo-EM density maps colored by local resolution in Å as calculated by RELION-3.0. (**D**) The models used in (B) that were built in each map colored according to local B-factor.

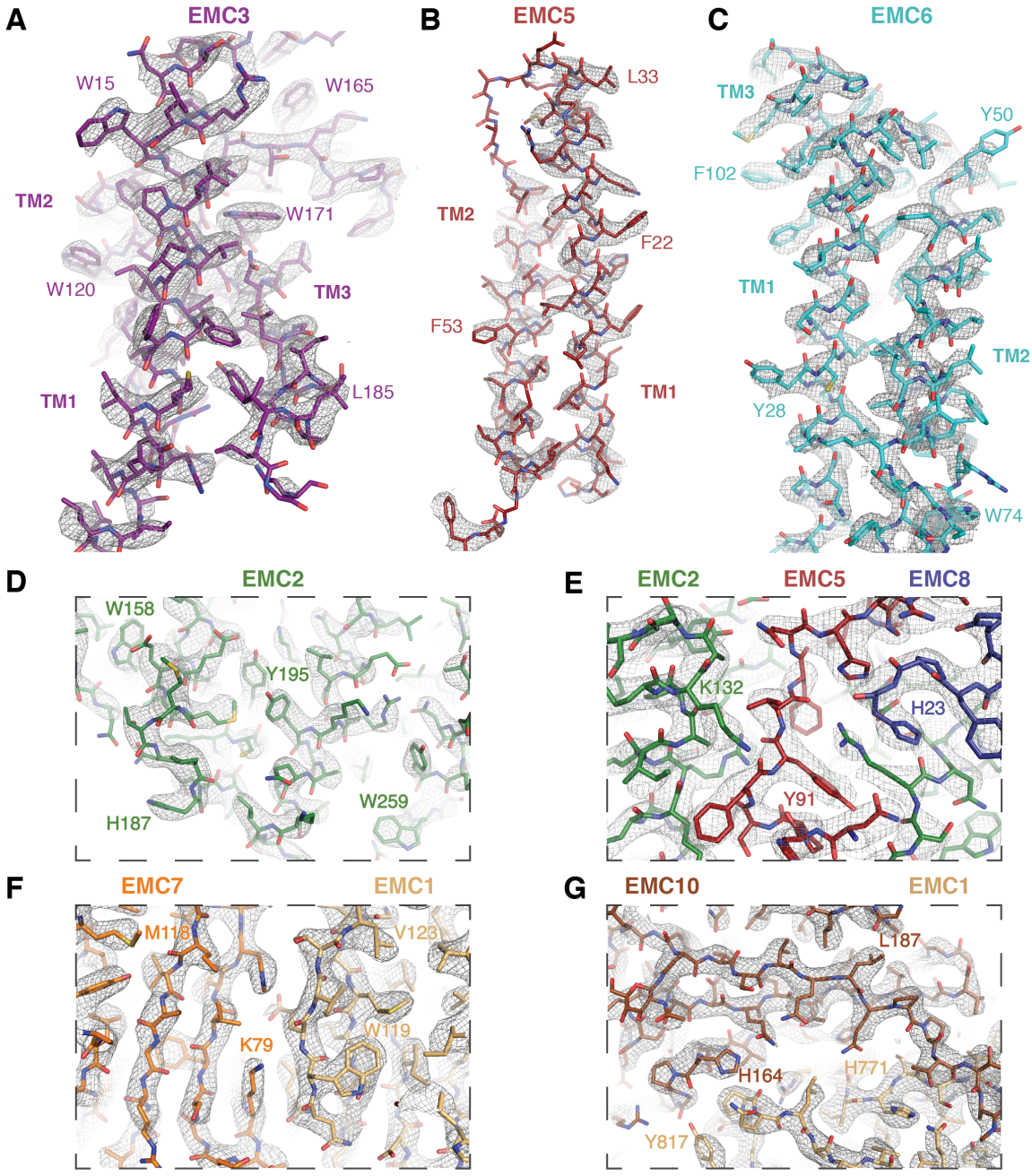
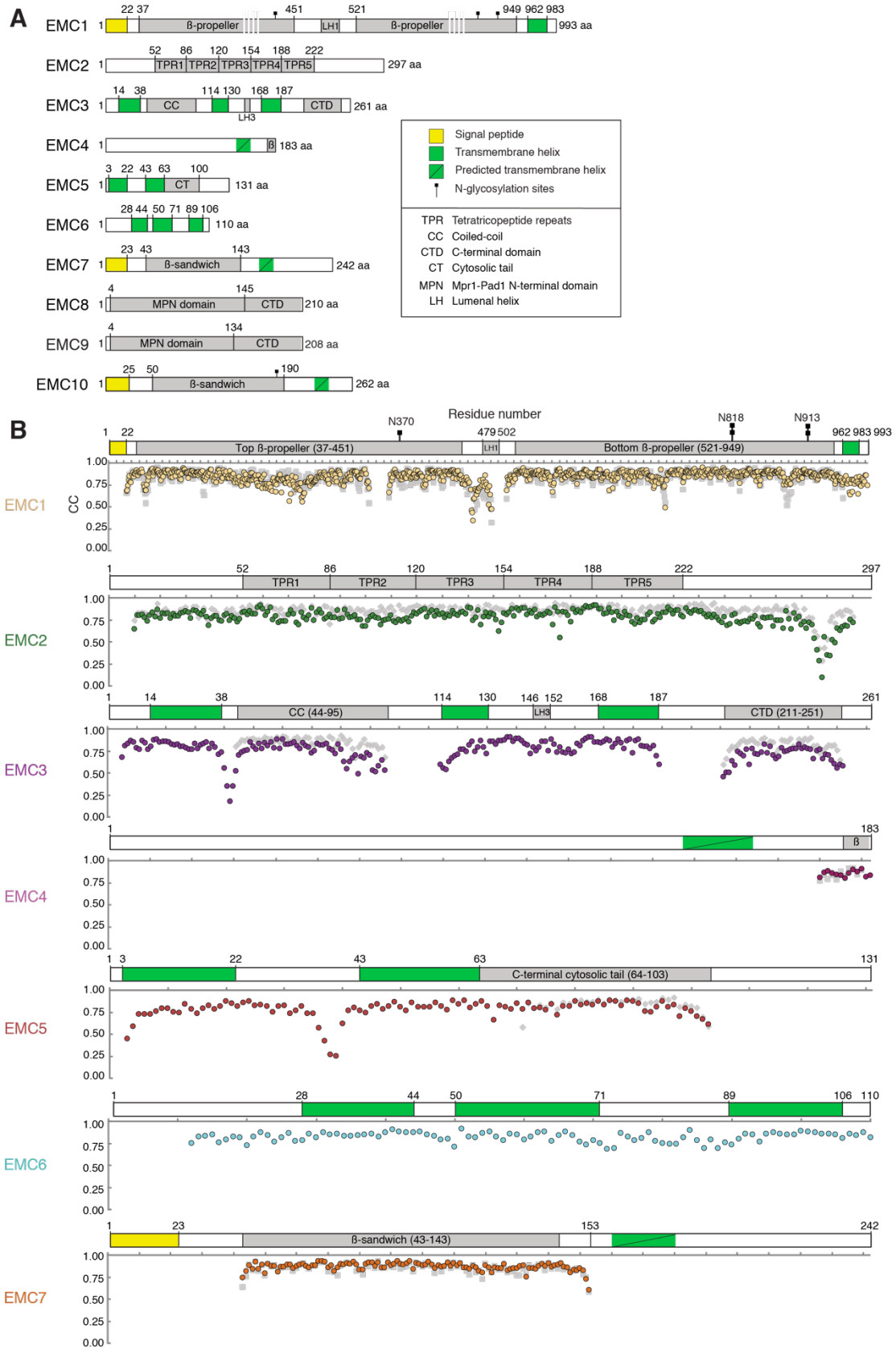


Fig. S4. Representative density for the transmembrane, cytosolic and lumenal regions.

(**A-C**) Representative density for the TMs of EMC3 (purple), 5 (red), and 6 (cyan), shown in the sharpened 'overall map' at 11, 9 and 10σ, respectively (fig. S2). (**D+E**) Representative density for the cytosolic region of the EMC, including EMC2 (green), the C-terminus of EMC5 (red), and EMC8 (blue) shown in the sharpened 'cytosolic map' at 16σ. (**F+G**) Representative density for the lumenal region of the EMC including EMC1 (gold), 7 (orange), and 10 (brown) shown in the sharpened 'lumenal map' at 8σ.


A
EMC1
EMC2
EMC3
EMC4
EMC5
EMC6
EMC7
EMC8
EMC9
EMC10
ß-propeller
LH1
TPR1
TPR2
TPR3
TPR4
TPR5
CC
CTD
LH3
CT
ß-sandwich
MPN domain
993 aa
297 aa
261 aa
183 aa
131 aa
110 aa
242 aa
210 aa
208 aa
262 aa
Signal peptide
Transmembrane helix
Predicted transmembrane helix
N-glycosylation sites
TPR Tetratricopeptide repeats
CC Coiled-coil
CTD C-terminal domain
CT Cytosolic tail
MPN Mpr1-Pad1 N-terminal domain
LH Lumenal helix
B
Residue number
N370
N818
N913
Top ß-propeller (37-451)
Bottom ß-propeller (521-949)
CC (44-95)
CTD (211-251)
C-terminal cytosolic tail (64-103)
ß-sandwich (43-143)
CC

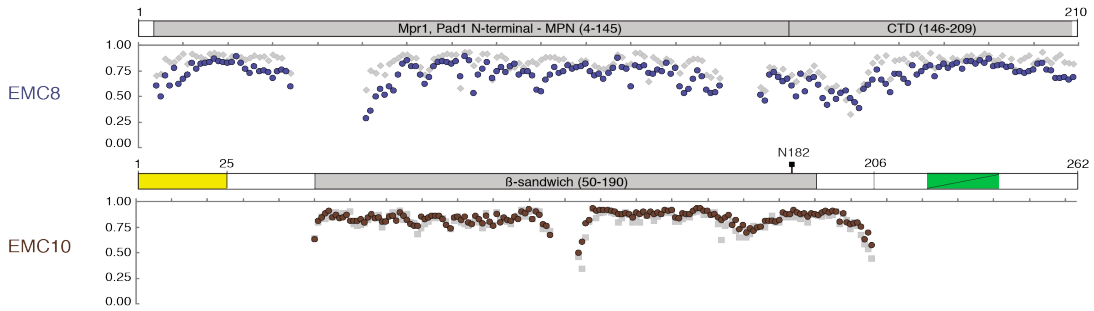


Fig. S5. Domain architecture and local properties of the individual domains of the EMC model.

(**A**) Schematic representation of the individual domain architecture of the EMC subunits. TPR domains in EMC2 were predicted using TPRpred (*60*) and validated by visual inspection. TPR domains are commonly found to mediate protein-protein interactions and in EMC2 serve to contact EMC8 and the cytosolic regions of EMC3, 5 and 1. EMC8 and EMC9 are vertebrate-specific paralogs resulting from a duplication event of an ancestral protein in the lineage leading to vertebrates (*9*). Binding of EMC8 to EMC2 is mutually exclusive to binding of EMC9. This is consistent with previous work showing that knockdown of EMC8 upregulates the cellular levels of EMC9 (*18*). EMC8 and 9 contain Mpr1- Pad1 N-terminal (MPN) domains, commonly found in quality control machinery, and could thus provide a potential link between the EMC and the ubiquitin proteasome pathway. The cytosolic domain of the EMC seems well-positioned to mediate interactions with the ribosome or other cytosolic co-factors of the EMC. Though the function of the EMC lumenal domain remains poorly defined, the β-propellers of EMC1 are homologous to other WD40-repeat proteins, including those found in the bacterial Bam insertase (*61*). These domains are known to function as scaffolds for protein-protein interactions and therefore may bind lumenal co-factors. Recent proteomics experiments suggest that the EMC is a platform for chaperone recruitment (*26*), and human mutations to surface exposed residues in EMC1 lead to developmental delays (*62*), potentially consistent with defects in protein folding or assembly. (**B**) Cross-correlation (CC) of the model for each subunit in the 'overall map' (colored circles), as well as the 'cytosolic' and 'lumenal' maps (grey diamonds and squares respectively), calculated in *phenix.real_space_refine*. The variation in CC values illustrates the non-uniform resolution of the map (fig. S3) and reflects the local quality of the model. It also highlights the improved local resolution of the 'cytosolic' and 'lumenal' maps. The density for the flexible linkers between the top ß-propeller and lumenal helix of EMC1 (amino acids 475-478), the α-helix 14 and α-helix 15 of EMC2 (amino acids 276-279), the TM1 and the coiled-coil of EMC3 (amino acids 40-43), and between the TM1 and TM2 of EMC5 (amino acids 37-39) was not sufficient to unambiguously assign the position of the side chains, thus they are included in the model as poly-Ala, yielding a local low cross-correlation value. EMC4 was initially built and refined as P170 to L183 (figs. S7, S8), yielding an overall cross-correlation of 0.79. However, fourteen residues are not sufficient to unambiguously assign the registry; therefore, in the model the region is assigned as poly-Ala/Gly.

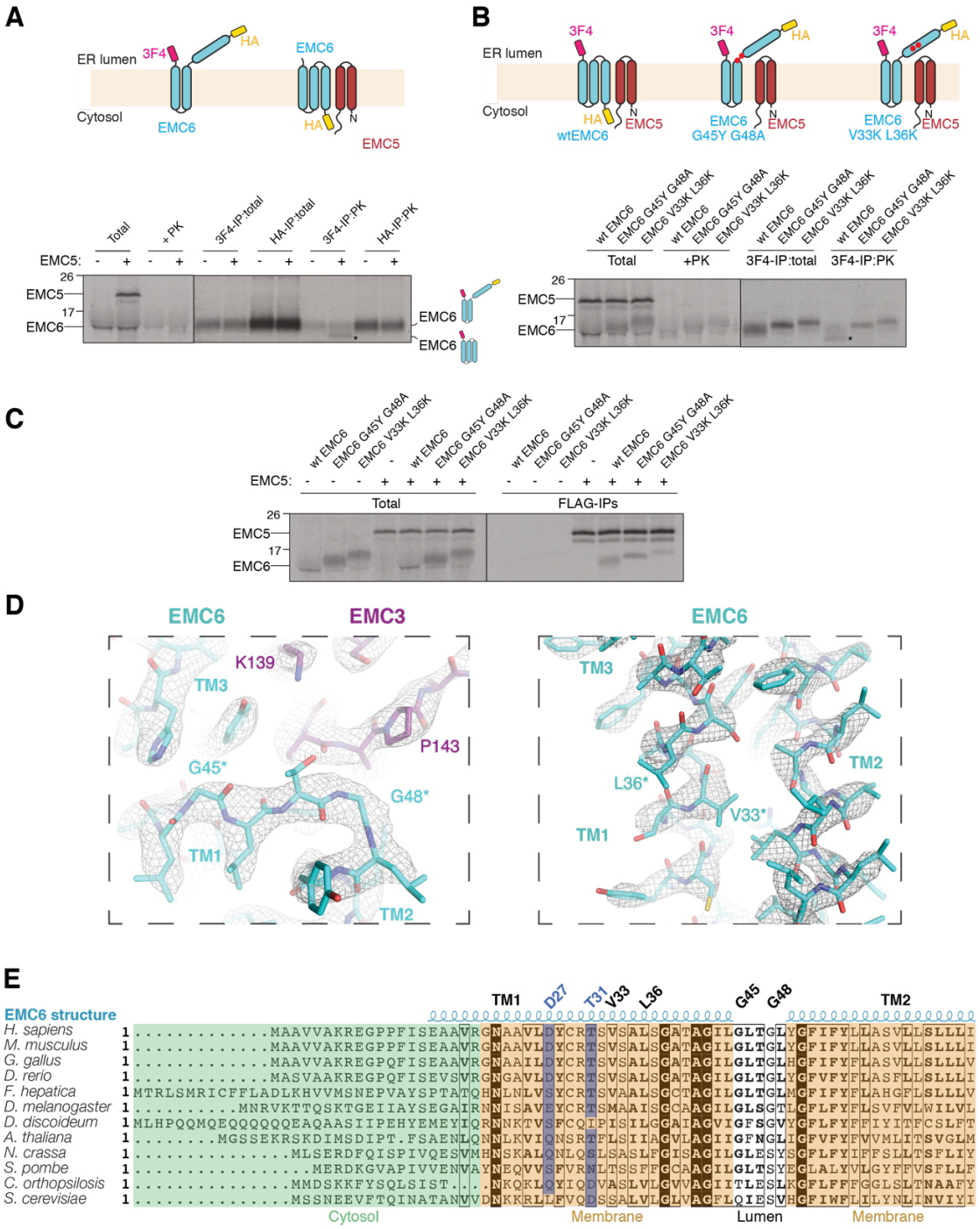


Fig. S6. The conserved EMC6 N-terminus forms a TM in the presence of EMC5.
(**A**) ^{35}S-methionine-labeled HA-EMC6-3F4 was translated in rabbit reticulocyte lysate (RRL) in the presence of canine-derived rough microsomes (cRMs) either alone or with EMC5-FLAG. The translations were treated with proteinase K (PK) and then analyzed directly, or following immunoprecipitation via the 3F4 or HA tag, by SDS-PAGE and autoradiography. Without EMC5, EMC6 inserts with the predicted topology, with two transmembrane helices and both termini in the lumen. The short loop between TMs 2 and

3 is protected from cleavage by the membrane and so the complete protein (with both tags) is recovered after PK treatment. Upon co-translation with EMC5, a second band appears (*) that is 3F4-tagged, but not HA-tagged. This represents an EMC6 species in which TM1 is now inserted into the bilayer resulting in the loss of the N-terminal HA-tag upon PK treatment. A schematic is shown above. (**B**) To validate this model, two conserved glycine residues in the TM1-2 loop (G45Y and G48A) or two hydrophobic residues in TM1 (V33K and L36K) were mutated and tested in the protease protection assay described in (A), in the presence of EMC5. In both cases we see a reduction in TM1 insertion compared to wild type (wt:*), suggesting both the flexibility of the TM1-2 loop and the hydrophobicity of TM1 are required for EMC5-dependent insertion. A schematic is shown above; the positions of the mutations are indicated with red circles. (**C**) To ascertain whether the inability of the EMC6 mutants to insert their first TM was due to inhibition of the interaction with EMC5, ^{35}S-methionine-labeled EMC6 variants (wild type, G45Y G48A, or V33K L36K) were translated in the presence or absence of EMC5-FLAG, and then FLAG-tagged complexes were immunoprecipitated. Both EMC6 mutants were recovered with EMC5-FLAG at similar levels to wild type. (**D**) Density for EMC6 TM1 and TM2 in the sharpened 'overall map' at 11σ. Asterisks indicate residues that were mutated in (B). (**E**) Sequence alignment of the N-terminus of EMC6 from higher to lower eukaryotes. The determined secondary structure of EMC6 for TM1 and TM2 is depicted above the alignment. Residues mutated within TM1 and the lumenal turn described in panel (B) are highlighted. Polar residues located in the hydrophilic vestibule and mutated in fig. S13D are highlighted in blue. Note that the highly conserved N22 in TM1 of EMC6 is within hydrogen-bonding distance of the main chain of EMC5. This is one of the many interactions of the new TM1 of EMC6 with EMC5 that may explain how this poor TM is stabilized in the bilayer (Table S3).

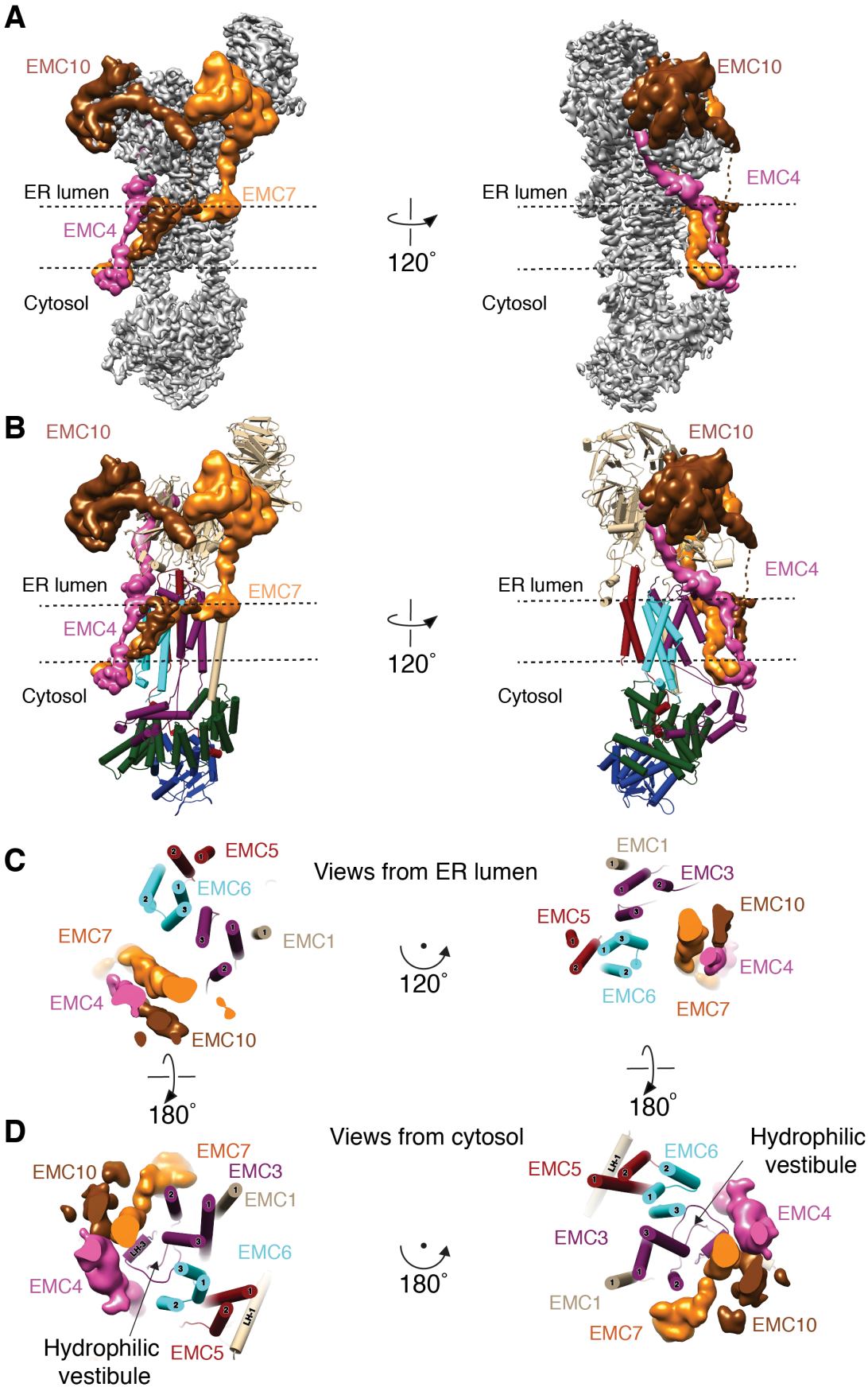

A
EMC10
ER lumen
EMC7
EMC4
Cytosol
120˚
EMC10
EMC4
ER lumen
Cytosol
B
EMC10
ER lumen
EMC7
EMC4
Cytosol
120˚
EMC10
EMC4
ER lumen
Cytosol
C
EMC5
EMC6
EMC7
EMC1
EMC4
EMC10
180˚
Views from ER lumen
EMC1
EMC3
EMC10
EMC5
EMC4
EMC6
EMC7
120˚
180˚
D
EMC7
EMC10
EMC3
EMC1
EMC6
EMC4
LH-3
LH-1
Hydrophilic
vestibule
EMC5
Views from cytosol
EMC6
EMC5
LH-1
Hydrophilic
vestibule
EMC4
EMC3
EMC1
EMC7
EMC10
180˚

Fig. S7. Density for the flexible TMs of EMC4, EMC7 and EMC10.
(**A**) Two views of the overall map (grey) at 6σ and a low-pass filtered map at 4σ, corresponding to EMC4 (magenta), EMC7 (orange), and EMC10 (brown) EM density regions. (**B**) Two views of the atomic model displayed as a cartoon superimposed with the low-pass filtered map for EMC4, EMC7, and EMC10. EM density consistent with the C-terminus of EMC4 was tentatively built as a poly-Ala/Gly model. Continuous density could be traced from this putative EMC4 β-strand, past the lumen-facing surface of EMC3, and through the lipid region of the nanodisc. We also detect strong EM density directly below the putative TM region of EMC4 that protrudes from the nanodisc/lipid and forms a surface with the cytosolic region of EMC2. It is likely that this additional density represents the N-terminal cytosolic region of EMC4. Therefore, though we cannot definitively assign the topology of EMC4, the structure is most consistent with a model in which it contains a single TM. Weak EM density for a TM is also detected C-terminal to the residue R152 of EMC7 (orange) and this density continues into the nanodisc lipid region. EM density that is C-terminal to L205 of EMC10 (brown) can be detected that continues toward the nanodisc/lipid at very low contour (and is depicted as a dashed line). Together the TMs of EMC4, 7, and 10 partially enclose the hydrophilic vestibule formed by EMC3 and 6. (**C**) Two views of the cross-section through the transmembrane region with the same coloring as (B), with a view from the lumen facing toward the cytosol. (**D**) Two views of the cross-section through the transmembrane region with a view from the cytosol facing up toward the membrane. When viewed from the cytosol, it is clear that substrate would encounter a hydrophilic vestibule that is capped by the lumenal helix of EMC3 (LH-3) and enveloped by the TMs of EMC4, 7, and 10.

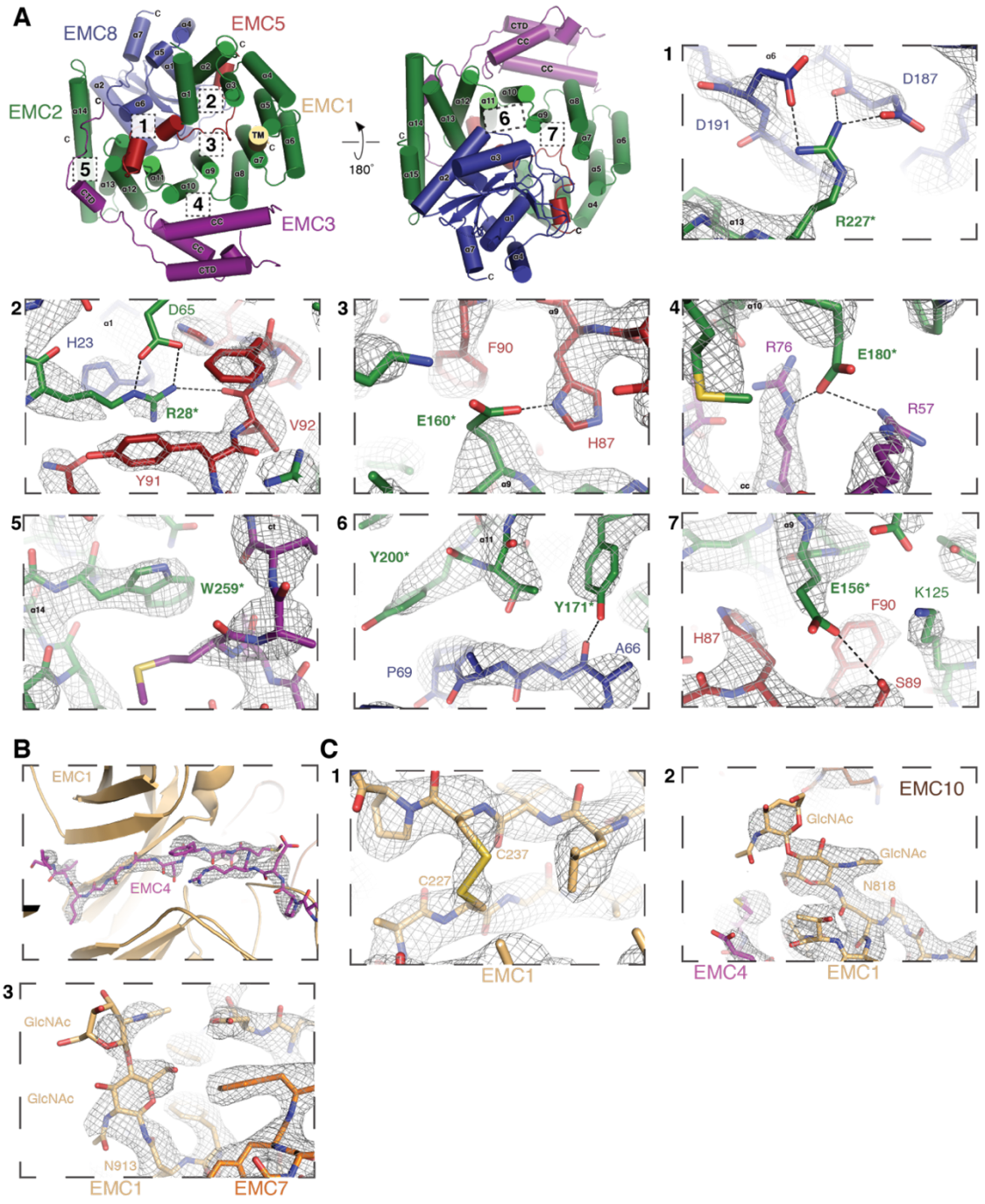


Fig. S8. Density for the cytosolic interactions and lumenal post-translational modifications.

(A) Top and bottom view of the cytosolic domain of the EMC. C = carboxyl-terminus, CTD = EMC3 C-terminal domain, CC = EMC3 coiled coil, α = alpha helix, TM = transmembrane helix**,** GlcNAc = N-acetylglucosamine. EMC2 serves as the scaffold for the cytosolic region of the EMC with five tetratricopeptide repeats (α3 - α12; fig. S5), which partly mediate protein-protein interactions with EMC8 and the cytosolic regions of EMC3, 5, and 1. Panels 1-7 display the density for the interaction surfaces between EMC2, 3, 5, and 8 in the cytosol. The views 1-4 correspond to those in Fig. 2B shown in

the sharpened ‘cytosolic map’ at 17σ (15σ in panel 5). Asterisks indicate residues in EMC2 that when mutated disrupt binding (Fig. 2C). (B) Close-up view of the density for the lumenal region of EMC4. EMC1 contains two eight-bladed β-propellers, one of which incorporates the single β-strand of EMC4 into a four-stranded blade (Fig. 2D, E). (C) ‘Lumenal map’ details. Panel 1 displays a disulfide bond in EMC1 (between residues 227 and 237), shown at 8σ. Panels 2-3 show N-linked acetylglucosamine (GlcNac) residues attached to N818 and N913 of EMC1 at the interface of EMC4, 7 and 10, shown at 7σ.

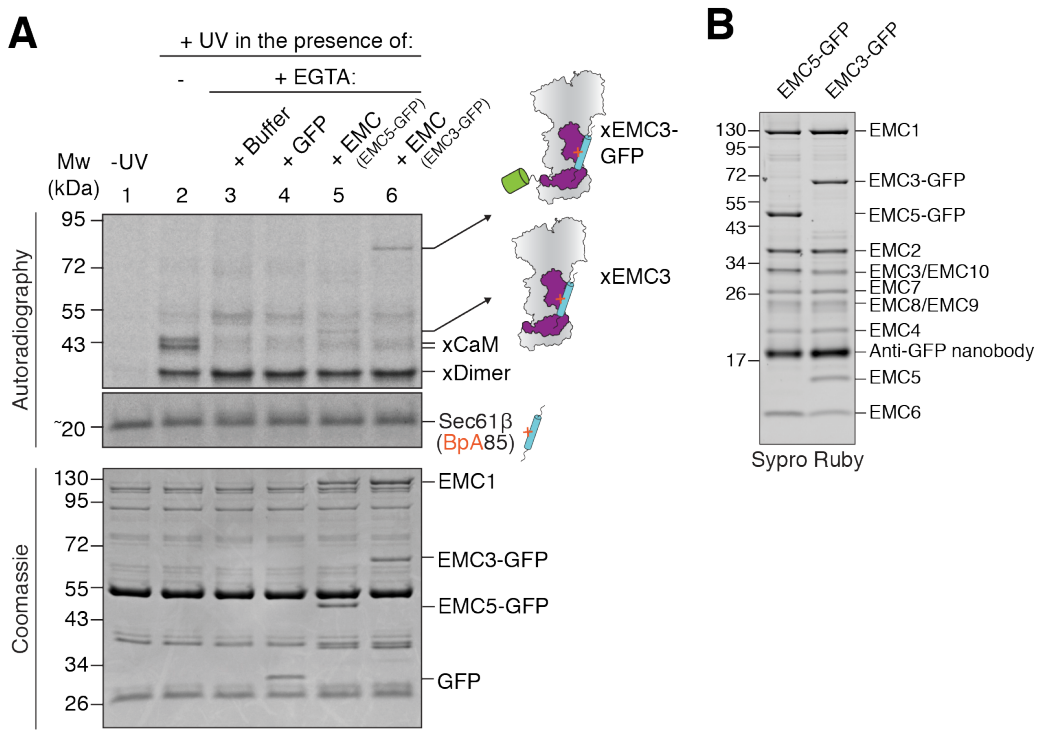


Fig. S9. EMC3 crosslinks to tail-anchored substrates.
(**A**) SEC61β, a known EMC tail-anchored substrate, was expressed with an amber stop codon at position 85 in its transmembrane helix in the presence of radioactive ^{35}S-methionine, 100 nM Ca^{2+} and 10 µM purified calmodulin (CaM) in the *E. coli*-derived PURE system in the absence of release factor RF1. The photoreactive unnatural amino acid 4-Benzoylphenylalanine (BpA) was incorporated at the amber stop codon using BpA synthetase and its supressor tRNA. After translation for 2 hours at 32˚C, 1 mM puromycin was added, the reaction was layered on top of a 20% (w/v) sucrose cushion prepared in physiologic salt buffer with 100 nM Ca^{2+}, and spun to remove aggregates. SEC61β-CaM complexes were retrieved from the cushion and incubated in the presence or absence of 1 mM EGTA, buffer (lane 3), GFP (lane 4), or detergent solubilized EMC affinity purified from either EMC5-GFP or EMC3-GFP cell lines (lanes 5 and 6 respectively). Except for the -UV control (lane 1), all reactions were irradiated with a UV source for 15 min on ice before quenching with SDS-PAGE sample buffer. Samples were analyzed by SDS-PAGE and autoradiography (top) or Coomassie staining (bottom). The ~20 kDa substrate produces a ~50 kDa crosslink in lane 5, consistent with crosslinking to ~30 kDa untagged EMC3. In lane 6, the EMC3-substrate crosslink shifts to ~80 kDa, consistent with addition of ~30 kDa from a linker-GFP fusion. Note that the absence of crosslinks to other EMC subunits does not exclude that these are also involved in directly contacting substrate transmembrane helices. (**B**) EMC complexes purified via EMC5- and EMC3-GFP as detailed in fig. S1 were analyzed by SDS-PAGE and Sypro Ruby staining.

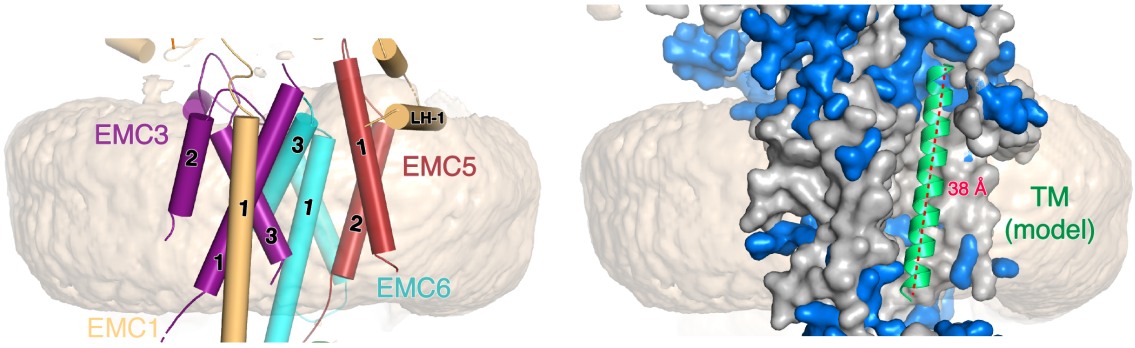


Fig. S10. The hydrophobic crevice can accommodate a full TM.
The 'hydrophobic crevice' shown in cartoon (left) and surface representation (right). Both views are depicted within a subtracted map of the nanodisc/lipid region to highlight the presence of membrane. Hydrophobic amino acids (in grey) are contributed by TMs 1 and 3 of EMC3, TMs 1 and 3 of EMC6, and TMs 1 and 2 of EMC5. Polar residues are shown in blue. The crevice is partially enclosed on the lumenal face by the lumenal helix of EMC1 (LH-1). A helix (lime) was modeled in the space-filling model to demonstrate that a 30 amino acid TM can be fully accommodated within the membrane spanning region of the hydrophobic crevice.

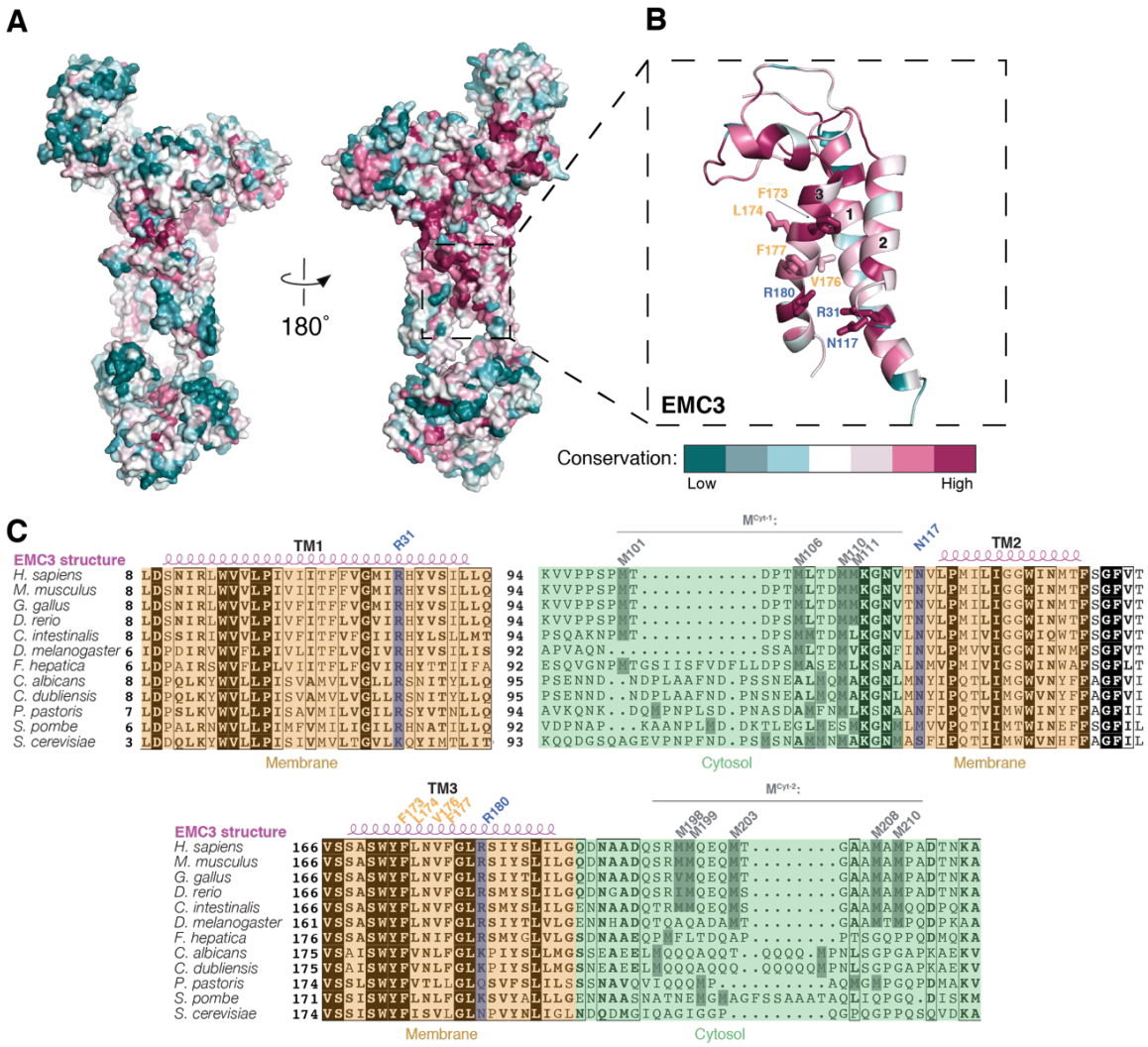


Fig. S11. Conservation of the hydrophilic vestibule and EMC3.
(**A**) A conservation plot mapped onto the structure of EMC to highlight the high conservation of the hydrophilic vestibule (right) as compared to the less conserved hydrophobic crevice (left). Many highly conserved residues on this side are located in EMC3 (boxed region), a member of the Oxa1 superfamily of insertases. (**B**) Conservation is mapped on to the cartoon representation of EMC3. Highly conserved residues are displayed as sticks and include the polar residues (labeled in blue) highlighted in Fig. 3B and mutated in Fig. 3C-E and fig. S13. (**C**) Sequence alignment of EMC3 that spans sequences derived from higher to lower eukaryotes. The panel is a sub-selection of species that were used for the conservation plot in panels A-B. The determined secondary structure of EMC3 is depicted above the alignment. The depicted regions are located in the membrane (TMs in tan) and cytosol (light green). Polar residues that were mutated are highlighted in blue. Conserved hydrophobic residues located above the polar patch are highlighted in gold. Methionine residues that comprise the 'Methionine-rich cytosolic loops' (M^{Cyt-1}: methionines 101, 106, 110, 111; M^{Cyt-2}: methionines 198, 199, 203, 208, 210) on the cytosolic side of TM2 and TM3 are highlighted in grey.

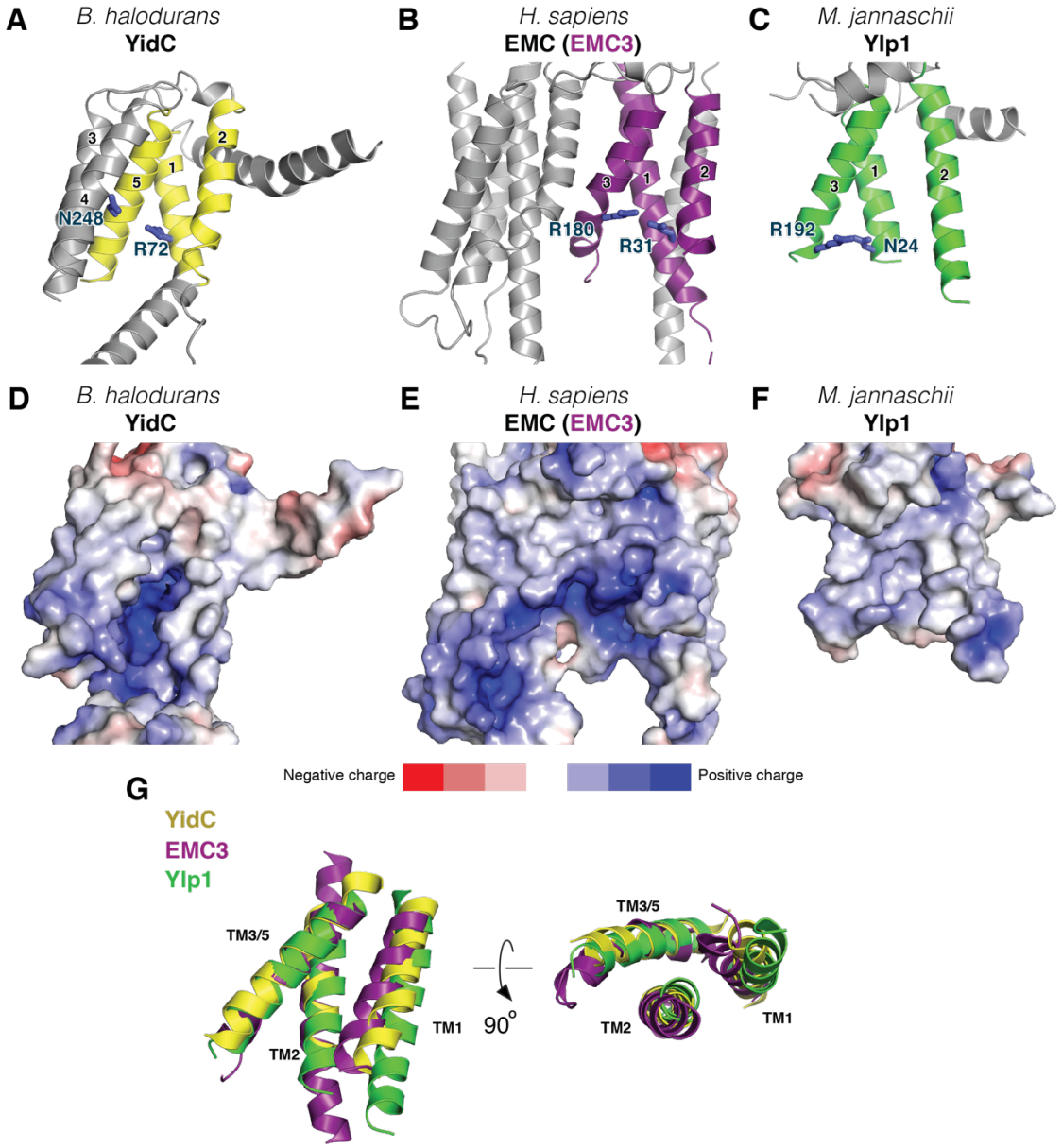


Fig. S12. Comparison of EMC3 to YidC and Ylp1.
(**A-C**) Secondary structure alignment of the putative insertase regions of YidC (TM1, TM2, TM5 in yellow, panel A), EMC3 (TM1, TM2, TM3 in deep purple, panel B), and Ylp1 (TM1, TM2, TM3 in green, panel C) in the context of their entire scaffold (grey). Polar residues implicated in the insertase function are depicted as blue sticks. (**D-F**) Electrostatic plots (generated using PDB2PQR and APBS with range -5 to +5 kT/e) mapped onto the surface of structures of *H. sapiens* EMC3, *B. halodurans* YidC (3WO6) (*19*), and *M. jannaschii* Ylp1 (5C8J) (*21*). Surface representations are shown in a similar orientation to panels A-C to highlight the positively charged vestibule. Seven polar residues can be found in EMC3 on the insertase side. (**G**) Orthogonal views of the superposition of all three insertases using the 'super' algorithm in PyMol.

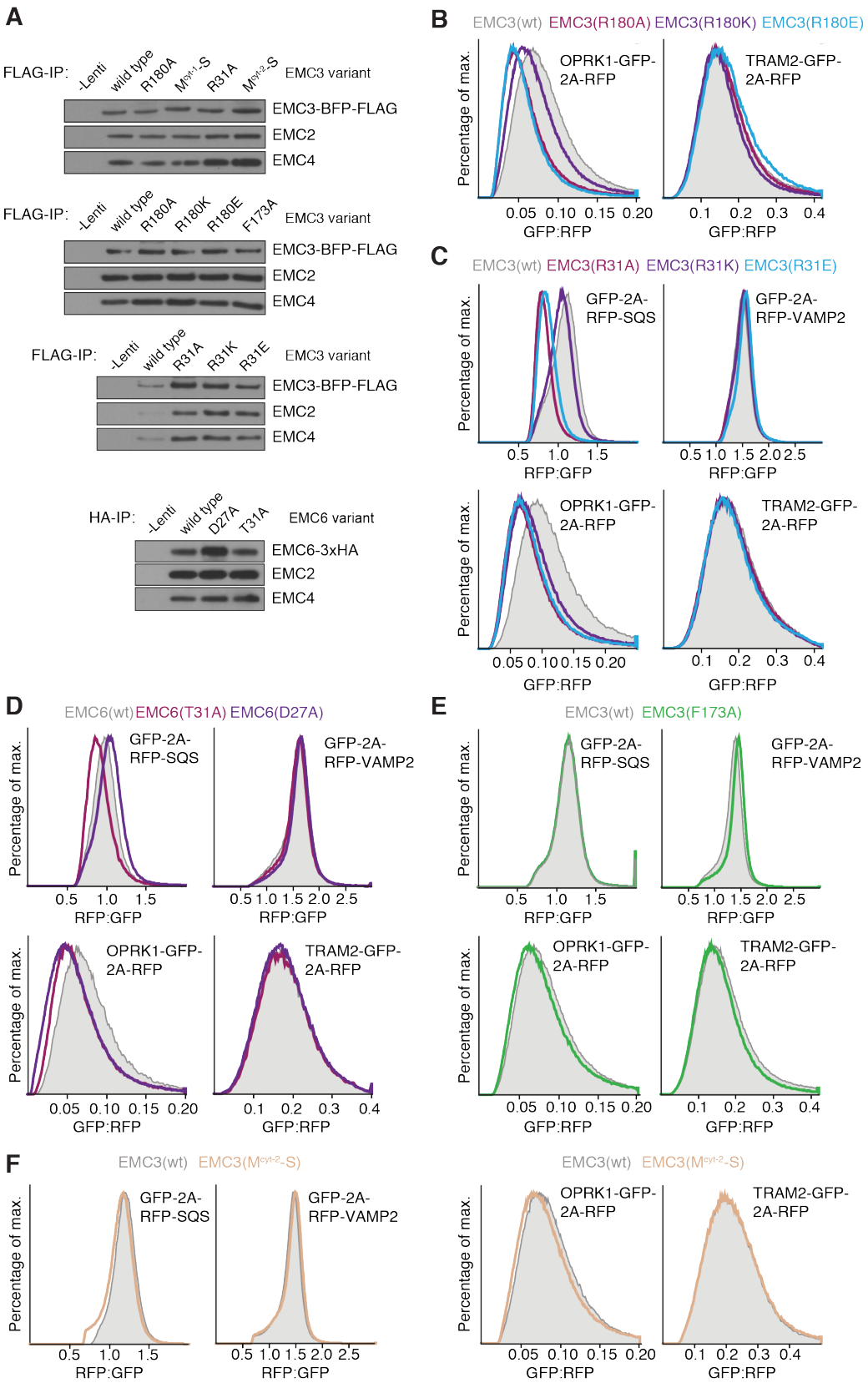

A
FLAG-IP:
-Lenti
wild type
R180A
M cyt-1-S
R31A
M cyt-2-S
EMC3 variant
EMC3-BFP-FLAG
EMC2
EMC4
FLAG-IP:
-Lenti
wild type
R180A
R180K
R180E
F173A
EMC3 variant
EMC3-BFP-FLAG
EMC2
EMC4
FLAG-IP:
-Lenti
wild type
R31A
R31K
R31E
EMC3 variant
EMC3-BFP-FLAG
EMC2
EMC4
HA-IP:
-Lenti
wild type
D27A
T31A
EMC6 variant
EMC6-3xHA
EMC2
EMC4
B
EMC3(wt) EMC3(R180A) EMC3(R180K) EMC3(R180E)
Percentage of max.
OPRK1-GFP-2A-RFP
TRAM2-GFP-2A-RFP
GFP:RFP
C
EMC3(wt) EMC3(R31A) EMC3(R31K) EMC3(R31E)
GFP-2A-RFP-SQS
GFP-2A-RFP-VAMP2
RFP:GFP
OPRK1-GFP-2A-RFP
TRAM2-GFP-2A-RFP
GFP:RFP
D
EMC6(wt) EMC6(T31A) EMC6(D27A)
GFP-2A-RFP-SQS
GFP-2A-RFP-VAMP2
RFP:GFP
OPRK1-GFP-2A-RFP
TRAM2-GFP-2A-RFP
GFP:RFP
E
EMC3(wt) EMC3(F173A)
GFP-2A-RFP-SQS
GFP-2A-RFP-VAMP2
RFP:GFP
OPRK1-GFP-2A-RFP
TRAM2-GFP-2A-RFP
GFP:RFP
F
EMC3(wt) EMC3(M cyt-2-S)
GFP-2A-RFP-SQS
GFP-2A-RFP-VAMP2
RFP:GFP
OPRK1-GFP-2A-RFP
TRAM2-GFP-2A-RFP
GFP:RFP

Fig. S13. Functional analysis of EMC3 and EMC6 mutations in the hydrophilic vestibule.

(**A**) EMC3 and EMC6 mutants incorporate efficiently into the EMC. HEK293 cells were transduced with lenti-viral vectors encoding either wild type or the indicated EMC3 mutants with a C-terminal TagBFP-3xFLAG tag. Wild type or mutant EMC6 was expressed with a C-terminal 3xHA-tag, followed by a viral 2A sequence and TagBFP. After 72 hours, cells were solubilized in detergent and subjected to FLAG or HA-IP. Bound proteins were eluted with 3xFLAG or 1xHA peptide and analyzed by SDS-PAGE and Western blotting with the indicated antibodies. Note that all shown EMC3 and EMC6 mutants co-purify soluble (EMC2) and other membrane-bound subunits (EMC4) to a similar extent as wild type. $M^{Cyt\text{-}1}$-S: mutation of methionines 101, 106, 110, 111 to serine; $M^{Cyt\text{-}2}$-S: mutation of methionines 198, 199, 203, 208, 210 to serine. These methionine residues were targeted because TM-binding chaperones, like SRP54 and Get3, are known to contain Met-rich substrate binding grooves (*22*). Though EMC2 could potentially contain a hydrophobic groove, in analogy to Get3, no Met-rich or otherwise hydrophobic structured binding site could be identified in the structure. In line with a previous observation (*21*), flexible Met-rich loops can be found adjacent to the TMs of EMC3. We found that Met-rich loop $M^{Cyt\text{-}1}$ was functionally important in cells (Fig. 3C), whereas $M^{Cyt\text{-}2}$ seemed to be dispensable (see F). (**B**) HEK293 cells stably expressing GFP-tagged OPRK1 (co-translational EMC substrate) or TRAM2 (EMC-independent multipass membrane protein) were transduced with lenti-viral vectors to stably express BFP-tagged wild type or mutant EMC3 to analyze the role of the positive charge on R180 in insertion. The GFP:RFP ratio of BFP-positive cells was analyzed by flow cytometry and then plotted as a histogram. (**C**) As in (B), testing the role of the positive charge on R31 in insertion. Additionally, cell lines expressing the post-translational EMC substrate SQS or the EMC-independent control protein VAMP2 were analyzed. (**D**) Assay as in (C), testing the effect of the two polar residues of EMC6 in the hydrophilic vestibule. (**E**) As in (C). No effect of the EMC3 F173A mutation on insertion of co- or post-translational substrates was observed. (**F**) As in (C). No effect of the methionine-rich loop $M^{Cyt\text{-}2}$-S mutant EMC3 on insertion of co- or post-translational substrates was observed.

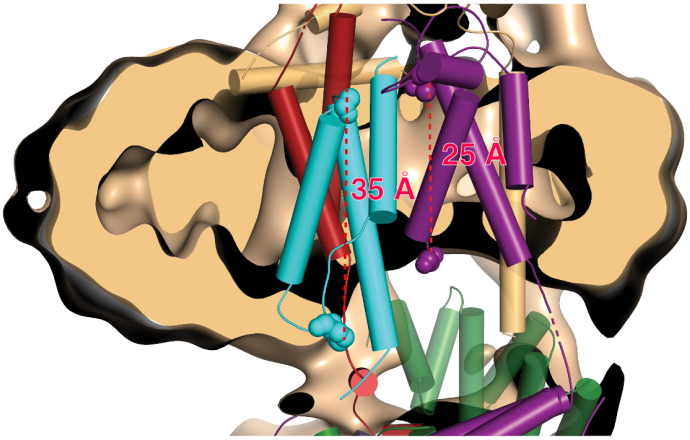


Fig. S14. Distance measurements of the lipid/nanodisc thickness.
Cartoon representation of the TM region of EMC with a cut-away view of a low-pass filtered, unsharpened overall EM density map (contoured at 3σ). Distances (in red) were measured from residues that are coincident with the boundaries of the lipid for the largest lipid width (EMC6 amino acids G48 to K79) and shortest width (EMC3 amino acids T138 to G188).

Table S1. Cryo-EM data collection, refinement and validation statistics

	Overall (EMDB-21929) (PDB 6WW7)	Lumenal (EMDB-21930)	Cytosolic (EMDB-21931)
Data collection and processing			
Microscope	FEI Titan Krios		
Voltage (kV)	300		
Camera	Gatan K3		
Energy filter	BioQuantum		
Energy filter slit width (eV)	20		
Magnification (nominal)	59,808 (130,000)		
Defocus range (μm)	-0.5 to -2.0		
Calibrated pixel size (Å/pix)	0.836		
Electron exposure (e^{-}/Å^2)	59.2		
Exposure rate (e^{-}/Å^2/frame)	1.48		
Number of frames per movie	40		
Automation software	SerialEM		
Number of micrographs	6345		
Initial particle images (no.)	1,034,250		
Final particle images (no.)	188,746	188,746	188,746
Estimated accuracy of translations (pix) (RELION)	0.87		
Estimated accuracy of rotations (°) (RELION)	1.571		
Local resolution range	2.5 – 6.5	3.0 – 6.0	2.5 – 4.0
Map resolution (Å, FSC=0.143)	3.4	3.2	3.6
Refinement			
Software (*phenix.real_space_refine*)	PHENIX 1.17.1-3660		
Initial model used (PDB code)	de novo	de novo	de novo
Resolution of unmasked reconstructions (Å, FSC= 0.5)	3.5	3.2	3.7
Resolution of masked reconstructions (Å, FSC= 0.5)	3.4	3.2	3.6
Correlation coefficient (CC_{mask})	0.83	0.87	0.84
Map sharpening *B* factor (Å^2)	-72	-97	-91
Model composition			
Non-hydrogen atoms	16576	9300	4872
Protein residues	2060	1167	592
Ligands	*UNK:14 NAG:6	NAG: 6	---
B factors (Å^2)	min/max/mean		
Protein	21/152/65	22/101/45	51/134/81
Ligand	23/89/54	47/66/56	---
R.M.S. deviations			
Bond lengths (Å) (# > 4σ)	0.003 (0)	0.004 (0)	0.004 (0)
Bond angles (°) (# > 4σ)	0.544(6)	0.622(2)	0.538 (0)
Validation			
MolProbity score	1.89	1.93	1.90
Clashscore	6.96	5.49	7.39
Poor rotamers (%)	2.92	3.62	2.7
Cß deviations (%)	0	0	0
CaBLAM outliers (%)	1.25	1.15	0.35
EMRinger score	2.50	3.65	2.91
Ramachandran plot			
Favored (%)	97.1	96.7	97.0
Allowed (%)	2.9	3.3	3.0
Disallowed (%)	0	0	0

*UNK is the code for the unknown amino acids of EMC4 assigned as poly-Ala/poly-Gly.

Table S2. Model building of the EMC

EMC	Chain	Region (residue range)	Region Name	Template for model building
1	A	24-36	NTD	de novo
		37-451	Top ß-propeller	trRosetta
		452-474	Coil/Helical	de novo
		479-502	LH1	de novo
		521-949	Bottom ß-propeller	trRosetta
		950-961	Coil	de novo
		962-983	TM	de novo
		984-993	CTD	de novo
2	B	10-51	NTD	de novo
		52-222	Core (TPR)	trRosetta
		223-290	CTD	de novo
3	C	5-13	NTD	de novo
		14-38	TM1	
		39-43	Coil	
		44-95	Coiled coil	
		114-130	TM2	
		131-167	Lumen region	
		168-187	TM3	
		211-252	CTD	
4	D	170-183	bottom ß-propeller	de novo (poly-Ala/poly-Gly)
		1-169	EMC4	sequence not modeled
5	E	3-22	TM1	de novo
		23-42	Lumen region - Helical	
		43-63	TM2	
		64-103	CT	
6	F	12-27	NTD - Helical	trRosetta / de novo
		28-44	TM1	trRosetta / de novo
		45-49	Lumen - Loop	de novo
		50-71	TM2	de novo
		72-88	Cytosol - Loop	de novo
		89-106	TM3	de novo
		107-110	CTD	de novo
7	G	43-143	ß-sandwich	trRosetta
		144-153	Coil	de novo
		154-242	TM / CTD	sequence not modeled
8	H	4-162	MPN	trRosetta
		163-209	CTD	de novo
10	I	50-190	ß-sandwich	trRosetta
		191-206	Coil	de novo
		207-262	TM / CTD	sequence not modeled

TM – transmembrane helix
LH-1 – lumenal helix of EMC1
NTD – amino-terminus domain
CTD – carboxyl-terminal domain
TPR – tetratricopeptide repeats
CT – cytosolic tail
MPN – Mpr1, Pad1 N-terminal

Table S3. Interface area between adjacent EMC subunits

Adjacent subunits of EMC	Buried area* (Å^2)
EMC 1 and EMC 2	386
EMC 1 and EMC 3	1595
EMC 1 and EMC 4	1139
EMC 1 and EMC 5	665
EMC 1 and EMC 7	1700
EMC 1 and EMC 10	1672
EMC 2 and EMC 3	1721
EMC 2 and EMC 5	1582
EMC 2 and EMC 8	1393
EMC 3 and EMC 5	76
EMC 3 and EMC 6	1053
EMC 3 and EMC 8	94
EMC 5 and EMC 6	1477
EMC 5 and EMC 8	634
EMC 7 and EMC 10	100

*Buried area indicates, in Å^2, the interface area calculated as difference in total accessible surface areas of isolated and interfacing structures divided by two (*29*, *30*).

Movie S1.

Movie depicting the structure of the human EMC, highlighting the cryo-EM density map and model shown in Figure 1.

Movie S2.

Movie highlighting the cryo-EM density for the dynamic transmembrane helices of EMC4, 7 and 10, which partially enclose the hydrophilic vestibule of the EMC.

Movie S3.

Movie of the cryo-EM density map and atomic model of the EMC depicting the local thinning of the lipid bilayer by the complex.

References and Notes

1. The UniProt Consortium, UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2017). doi:10.1093/nar/gkw1099 Medline

2. N. Aviram, T. Ast, E. A. Costa, E. C. Arakel, S. G. Chuartzman, C. H. Jan, S. Haßdenteufel, J. Dudek, M. Jung, S. Schorr, R. Zimmermann, B. Schwappach, J. S. Weissman, M. Schuldiner, The SND proteins constitute an alternative targeting route to the endoplasmic reticulum. *Nature* **540**, 134–138 (2016). doi:10.1038/nature20169 Medline

3. A. Guna, N. Volkmar, J. C. Christianson, R. S. Hegde, The ER membrane protein complex is a transmembrane domain insertase. *Science* **359**, 470–473 (2018). doi:10.1126/science.aao3099 Medline

4. D. Görlich, T. A. Rapoport, Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane. *Cell* **75**, 615–630 (1993). doi:10.1016/0092-8674(93)90483-7 Medline

5. M. Mariappan, A. Mateja, M. Dobosz, E. Bove, R. S. Hegde, R. J. Keenan, The mechanism of membrane-associated steps in tail-anchored protein insertion. *Nature* **477**, 61–66 (2011). doi:10.1038/nature10362 Medline

6. M. Schuldiner, J. Metz, V. Schmid, V. Denic, M. Rakwalska, H. D. Schmitt, B. Schwappach, J. S. Weissman, The GET complex mediates insertion of tail-anchored proteins into the ER membrane. *Cell* **134**, 634–645 (2008). doi:10.1016/j.cell.2008.06.025 Medline

7. S. Stefanovic, R. S. Hegde, Identification of a targeting factor for posttranslational membrane protein insertion into the ER. *Cell* **128**, 1147–1159 (2007). doi:10.1016/j.cell.2007.01.036 Medline

8. M. C. Jonikas, S. R. Collins, V. Denic, E. Oh, E. M. Quan, V. Schmid, J. Weibezahn, B. Schwappach, P. Walter, J. S. Weissman, M. Schuldiner, Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science* **323**, 1693–1697 (2009). doi:10.1126/science.1167983 Medline

9. J. G. Wideman, The ubiquitous and ancient ER membrane protein complex (EMC): Tether or not? *F1000Research* **4**, 624 (2015). doi:10.12688/f1000research.6944.1 Medline

10. P. J. Chitwood, R. S. Hegde, The Role of EMC during Membrane Protein Biogenesis. *Trends Cell Biol.* **29**, 371–384 (2019). doi:10.1016/j.tcb.2019.01.007 Medline

11. P. J. Chitwood, S. Juszkiewicz, A. Guna, S. Shao, R. S. Hegde, EMC Is Required to Initiate Accurate Membrane Protein Topogenesis. *Cell* **175**, 1507–1519.e16 (2018). doi:10.1016/j.cell.2018.10.009 Medline

12. S. W. Hennon, R. Soman, L. Zhu, R. E. Dalbey, YidC/Alb3/Oxa1 Family of Insertases. *J. Biol. Chem.* **290**, 14866–14874 (2015). doi:10.1074/jbc.R115.638171 Medline

13. S. A. Anghel, P. T. McGilvray, R. S. Hegde, R. J. Keenan, Identification of Oxa1 Homologs Operating in the Eukaryotic Endoplasmic Reticulum. *Cell Rep.* **21**, 3708–3716 (2017). doi:10.1016/j.celrep.2017.12.006 Medline

14. T. Pleiner, M. Bates, S. Trakhanov, C.-T. Lee, J. E. Schliep, H. Chug, M. Böhning, H. Stark, H. Urlaub, D. Görlich, Nanobodies: Site-specific labeling for super-resolution imaging,

rapid epitope-mapping and native protein complex isolation. *eLife* **4**, e11349 (2015). doi:10.7554/eLife.11349 Medline

15. T. Hessa, N. M. Meindl-Beinker, A. Bernsel, H. Kim, Y. Sato, M. Lerch-Bader, I. Nilsson, S. H. White, G. von Heijne, Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**, 1026–1030 (2007). doi:10.1038/nature06387 Medline

16. H. J. F. Carvalho, A. Del Bondio, F. Maltecca, S. F. Colombo, N. Borgese, The WRB Subunit of the Get3 Receptor is Required for the Correct Integration of its Partner CAML into the ER. *Sci. Rep.* **9**, 11887 (2019). doi:10.1038/s41598-019-48363-2 Medline

17. A. J. Inglis, K. R. Page, A. Guna, R. M. Voorhees, Differential Modes of Orphan Subunit Recognition for the WRB/CAML Complex. *Cell Rep.* **30**, 3691–3698.e5 (2020). doi:10.1016/j.celrep.2020.02.084 Medline

18. N. Volkmar, M.-L. Thezenas, S. M. Louie, S. Juszkiewicz, D. K. Nomura, R. S. Hegde, B. M. Kessler, J. C. Christianson, The ER membrane protein complex promotes biogenesis of sterol-related enzymes maintaining cholesterol homeostasis. *J. Cell Sci.* **132**, jcs223453 (2019). doi:10.1242/jcs.223453 Medline

19. K. Kumazaki, S. Chiba, M. Takemoto, A. Furukawa, K. Nishiyama, Y. Sugano, T. Mori, N. Dohmae, K. Hirata, Y. Nakada-Nakura, A. D. Maturana, Y. Tanaka, H. Mori, Y. Sugita, F. Arisaka, K. Ito, R. Ishitani, T. Tsukazaki, O. Nureki, Structural basis of Sec-independent membrane protein insertion by YidC. *Nature* **509**, 516–520 (2014). doi:10.1038/nature13167 Medline

20. G. Drin, B. Antonny, Amphipathic helices and membrane curvature. *FEBS Lett.* **584**, 1840–1847 (2010). doi:10.1016/j.febslet.2009.10.022 Medline

21. M. T. Borowska, P. K. Dominik, S. A. Anghel, A. A. Kossiakoff, R. J. Keenan, A YidC-like Protein in the Archaeal Plasma Membrane. *Structure* **23**, 1715–1724 (2015). doi:10.1016/j.str.2015.06.025 Medline

22. A. Guna, R. S. Hegde, Transmembrane Domain Recognition during Membrane Protein Biogenesis and Quality Control. *Curr. Biol.* **28**, R498–R511 (2018). doi:10.1016/j.cub.2018.02.004 Medline

23. K. Mitra, I. Ubarretxena-Belandia, T. Taguchi, G. Warren, D. M. Engelman, Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 4083–4088 (2004). doi:10.1073/pnas.0307332101 Medline

24. Y. Chen, S. Capponi, L. Zhu, P. Gellenbeck, J. A. Freites, S. H. White, R. E. Dalbey, YidC Insertase of *Escherichia coli*: Water Accessibility and Membrane Shaping. *Structure* **25**, 1403–1414.e3 (2017). doi:10.1016/j.str.2017.07.008 Medline

25. X. Wu, M. Siggel, S. Ovchinnikov, W. Mi, V. Svetlov, E. Nudler, M. Liao, G. Hummer, T. A. Rapoport, Structural basis of ER-associated protein degradation mediated by the Hrd1 ubiquitin ligase complex. *Science* **368**, eaaz2449 (2020). doi:10.1126/science.aaz2449 Medline

26. M. J. Shurtleff, D. N. Itzhak, J. A. Hussmann, N. T. Schirle Oakdale, E. A. Costa, M. Jonikas, J. Weibezahn, K. D. Popova, C. H. Jan, P. Sinitcyn, S. S. Vembar, H.

Hernandez, J. Cox, A. L. Burlingame, J. L. Brodsky, A. Frost, G. H. H. Borner, J. S. Weissman, The ER membrane protein complex interacts cotranslationally to enable biogenesis of multipass membrane proteins. *eLife* **7**, e37018 (2018). doi:10.7554/eLife.37018 Medline

27. S. Tian, Q. Wu, B. Zhou, M. Y. Choi, B. Ding, W. Yang, M. Dong, Proteomic Analysis Identifies Membrane Proteins Dependent on the ER Membrane Protein Complex. *Cell Rep.* **28**, 2517–2526.e5 (2019). doi:10.1016/j.celrep.2019.08.006 Medline

28. R. M. Voorhees, I. S. Fernández, S. H. Scheres, R. S. Hegde, Structure of the mammalian ribosome-Sec61 complex to 3.4 Å resolution. *Cell* **157**, 1632–1643 (2014). doi:10.1016/j.cell.2014.05.024 Medline

29. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007). doi:10.1016/j.jmb.2007.05.022 Medline

30. R. A. Laskowski, J. Jabłońska, L. Pravda, R. S. Vařeková, J. M. Thornton, PDBsum: Structural summaries of PDB entries. *Protein Sci.* **27**, 129–134 (2018). doi:10.1002/pro.3289 Medline

31. S. Frey, D. Görlich, A new set of highly efficient, tag-cleaving proteases for purifying recombinant proteins. *J. Chromatogr. A* **1337**, 95–105 (2014). doi:10.1016/j.chroma.2014.02.029 Medline

32. T. Pleiner, M. Bates, D. Görlich, A toolbox of anti-mouse and anti-rabbit IgG secondary nanobodies. *J. Cell Biol.* **217**, 1143–1154 (2018). doi:10.1083/jcb.201709115 Medline

33. A. Vera Rodriguez, S. Frey, D. Görlich, Engineered SUMO/protease system identifies Pdr6 as a bidirectional nuclear transport receptor. *J. Cell Biol.* **218**, 2006–2020 (2019). doi:10.1083/jcb.201812091 Medline

34. M. Mariappan, X. Li, S. Stefanovic, A. Sharma, A. Mateja, R. J. Keenan, R. S. Hegde, A ribosome-associating factor chaperones tail-anchored membrane proteins. *Nature* **466**, 1120–1124 (2010). doi:10.1038/nature09296 Medline

35. S. Shao, M. C. Rodrigo-Brenni, M. H. Kivlen, R. S. Hegde, Mechanistic basis for a molecular triage reaction. *Science* **355**, 298–302 (2017). doi:10.1126/science.aah6130 Medline

36. T. K. Ritchie, Y. V. Grinkova, T. H. Bayburt, I. G. Denisov, J. K. Zolnerciks, W. M. Atkins, S. G. Sligar, Chapter 11 - Reconstitution of membrane proteins in phospholipid bilayer nanodiscs. *Methods Enzymol.* **464**, 211–231 (2009). doi:10.1016/S0076-6879(09)64011-8 Medline

37. A. Kirchhofer, J. Helma, K. Schmidthals, C. Frauer, S. Cui, A. Karcher, M. Pellis, S. Muyldermans, C. S. Casas-Delucchi, M. C. Cardoso, H. Leonhardt, K.-P. Hopfner, U. Rothbauer, Modulation of protein properties in living cells using nanobodies. *Nat. Struct. Mol. Biol.* **17**, 133–138 (2010). doi:10.1038/nsmb.1727 Medline

38. D. Beckett, E. Kovaleva, P. J. Schatz, A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* **8**, 921–929 (1999). doi:10.1110/ps.8.4.921 Medline

39. M. Fairhead, M. Howarth, Site-specific biotinylation of purified proteins using BirA. *Methods Mol. Biol.* **1266**, 171–184 (2015). doi:10.1007/978-1-4939-2272-7_12 Medline

40. L. Liu, J. Spurrier, T. R. Butt, J. E. Strickler, Enhanced protein expression in the baculovirus/insect cell system using engineered SUMO fusions. *Protein Expr. Purif.* **62**, 21–28 (2008). doi:10.1016/j.pep.2008.07.010 Medline

41. A. Sharma, M. Mariappan, S. Appathurai, R. S. Hegde, In vitro dissection of protein translocation into the mammalian endoplasmic reticulum. *Methods Mol. Biol.* **619**, 339–363 (2010). doi:10.1007/978-1-60327-412-8_20 Medline

42. P. Walter, G. Blobel, Preparation of microsomal membranes for cotranslational protein translocation. *Methods Enzymol.* **96**, 84–93 (1983). doi:10.1016/S0076-6879(83)96010-X Medline

43. E. Itakura, E. Zavodszky, S. Shao, M. L. Wohlever, R. J. Keenan, R. S. Hegde, Ubiquilins Chaperone and Triage Mitochondrial Membrane Proteins for Degradation. *Mol. Cell* **63**, 21–33 (2016). doi:10.1016/j.molcel.2016.05.020 Medline

44. P. de Felipe, G. A. Luke, L. E. Hughes, D. Gani, C. Halpin, M. D. Ryan, E unum pluribus: Multiple proteins from a self-processing polyprotein. *Trends Biotechnol.* **24**, 68–75 (2006). doi:10.1016/j.tibtech.2005.12.006 Medline

45. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005). doi:10.1016/j.jsb.2005.07.007 Medline

46. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017). doi:10.1038/nmeth.4169 Medline

47. J. Zivanov, T. Nakane, B. O. Forsberg, D. Kimanius, W. J. H. Hagen, E. Lindahl, S. H. W. Scheres, New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018). doi:10.7554/eLife.42166 Medline

48. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). doi:10.1002/jcc.20084 Medline

49. T. C. Terwilliger, P. D. Adams, P. V. Afonine, O. V. Sobolev, A fully automatic method yielding initial models from high-resolution cryo-electron microscopy maps. *Nat. Methods* **15**, 905–908 (2018). doi:10.1038/s41592-018-0173-1 Medline

50. A. Casañal, B. Lohkamp, P. Emsley, Current developments in Coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. *Protein Sci.* **29**, 1069–1078 (2020). doi:10.1002/pro.3791 Medline

51. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010). doi:10.1107/S0907444910007493 Medline

52. C. M. Ho, X. Li, M. Lai, T. C. Terwilliger, J. R. Beck, J. Wohlschlegel, D. E. Goldberg, A. W. P. Fitzpatrick, Z. H. Zhou, Bottom-up structural proteomics: cryoEM of protein complexes enriched from the cellular milieu. *Nat. Methods* **17**, 79–85 (2020). doi:10.1038/s41592-019-0637-y Medline

53. P. V. Afonine, B. K. Poon, R. J. Read, O. V. Sobolev, T. C. Terwilliger, A. Urzhumtsev, P. D. Adams, Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018). doi:10.1107/S2059798318006551 Medline

54. J. Yang, I. Anishchenko, H. Park, Z. Peng, S. Ovchinnikov, D. Baker, Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 1496–1503 (2020). doi:10.1073/pnas.1914677117 Medline

55. D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L.-W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D. Adams, Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. D* **75**, 861–877 (2019). doi:10.1107/S2059798319011471 Medline

56. D. T. Jones, Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999). doi:10.1006/jmbi.1999.3091 Medline

57. T. I. Croll, ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D* **74**, 519–530 (2018). doi:10.1107/S2059798318002425 Medline

58. N. W. Moriarty, R. W. Grosse-Kunstleve, P. D. Adams, electronic Ligand Builder and Optimization Workbench (eLBOW): A tool for ligand coordinate and restraint generation. *Acta Crystallogr. D* **65**, 1074–1080 (2009). doi:10.1107/S0907444909029436 Medline

59. V. B. Chen, W. B. Arendall 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010). doi:10.1107/S0907444909042073 Medline

60. M. R. Karpenahalli, A. N. Lupas, J. Söding, TPRpred: A tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences. *BMC Bioinformatics* **8**, 2 (2007). doi:10.1186/1471-2105-8-2 Medline

61. N. Noinaj, J. C. Gumbart, S. K. Buchanan, The β-barrel assembly machinery in motion. *Nat. Rev. Microbiol.* **15**, 197–204 (2017). doi:10.1038/nrmicro.2016.191 Medline

62. T. Harel, G. Yesil, Y. Bayram, Z. Coban-Akdemir, W.-L. Charng, E. Karaca, A. Al Asmari, M. K. Eldomery, J. V. Hunter, S. N. Jhangiani, J. A. Rosenfeld, D. Pehlivan, A. W. El-Hattab, M. A. Saleh, C. A. LeDuc, D. Muzny, E. Boerwinkle, R. A. Gibbs, W. K. Chung, Y. Yang, J. W. Belmont, J. R. Lupski; Baylor-Hopkins Center for Mendelian Genomics, Monoallelic and Biallelic Variants in EMC1 Identified in Individuals with Global Developmental Delay, Hypotonia, Scoliosis, and Cerebellar Atrophy. *Am. J. Hum. Genet.* **98**, 562–570 (2016). doi:10.1016/j.ajhg.2016.01.011 Medline